EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Experimental Methods and Simulation Techniques: What can be learned about trust, schooling decisions and exchange rates?

Fabian Bornhorst

Thesis submitted for assessment with a view to obtaining the degree of Doctor of the European University Institute

Florence
September 2004

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Experimental Methods and Simulation Techniques: What can be learned about trust, schooling decisions and exchange rates?

Fabian Bornhorst

The Thesis Committee consists of:

Prof. Anindya Banerjee, supervisor, EUI
" Alan Duncan, University of Nottingham
" Urs Fischbacher, University of Zurich
" Andrea Ichino, EUI

Acknowledgements

I want to thank my supervisors Anindya Banerjee, Andrea Ichino, and Karl Schlag for their continuous support, encouragement and valuable advice throughout my time at the European University Institute (EUI). I also wish to thank my co-authors Andrea Ichino, Karl Schlag and Eyal Winter for our joint work on the experiment, which is analyzed in the first part of this thesis. My gains from participating in this project go far beyond the chapters included in this dissertation.

Every chapter of this thesis benefitted enormously from comments received in seminars and discussions both in the EUI and outside. In particular, I wish to thank Christopher Baum, Francisco Ferreira, Urs Fischbacher, Oliver Kirchkamp, Phillippe Leite, Helmut Lütkepohl, Daniel McFadden, Maite Martínez-Granado and Frank Vella for critical and insightful comments. I also thank all members of the interdisciplinary research group *Social and Economic Networks* at the EUI who were involved in the experiment. My colleagues and friends Jürgen Maurer, Gerry Müller and Roland Straub, besides from being a good and reliable company in *La Pantera Viola*, also gave me very helpful comments at various stages of this work.

Without the efficient and friendly support from the secretaries of the Economics department, especially Jessica Spataro, it would be hard to complete a dissertation. Thanks for all your work!

I gratefully acknowledge financial support from the German Academic Exchange Service (DAAD) and from the EUI. The Research Council of the EUI kindly provided the funds to run the experiment.

Finally, I owe a big thanks to all my friends inside and outside the academic world for making these four years such an enjoyable and exciting period of my life. My exposure to social scientists other than economists at the institute was healthy in many respects. I especially thank Carsten Q. Schneider for his support in countless circumstances and also Maria Jiménez Buedo and Sílvia Conduto de Sousa for making me feel at home in *Orsini 38*. Thanks to my friends Bergljot Barkbu and Emily Sinnott even the most depressing day while writing this dissertation could turn into an unforgettable Florence experience.

I am indebted to my family for their wide ranging support and their constant encouragement. Last but not least I want to thank Hannah Nielsen for her detailed comments on the thesis and, more than anything, for being there for me. I cannot imagine myself where I am today without her support.

Florence, September 2001 FABIAN BORNHORST

Contents

Introduction

Economic theory is an attempt to explain motives and consequences of individual actions by modelling the interaction of economic agents and deducing theories that relate economic variables at the aggregate level. This approach has contributed significantly to the understanding of economic issues and forms the basis for economic policy. Yet economic research can only make useful contributions as a social science if it does not shy away from confronting its theoretical predictions with observable phenomena.

Rationality is the predominant behavioural principle of economic theory. With few exceptions, economic models build on agents that behave in a rational way. But in its purest form rationality fails to describe, let alone explain, several aspects of human behaviour. Observable characteristics such as trust or fairness cannot be explained within the standard framework of individual utility maximization. While there is no doubt that rationality remains a useful benchmark for describing behaviour in many circumstances, the undeniable existence of something beyond its reach not only challenges positive economics but also puts into question predictions of standard economic theory.

Experimental techniques are increasingly being used in economics in order to investigate the motives and economic consequences of individual actions. They have proven useful in the laboratory, mainly to investigate the foundations of individual behaviour in a controlled environment. But also in the field experimental methods are advantageous, particularly to identify the effect of economic policies by comparing the treatment with a control group.

In general, the increased availability of economic data, both at the micro- and at the macroeconomic level, has lead to a surge in empirical testing of economic hypotheses. Parallel to that, rapid advances in computing technology have made quantitative techniques available to a wide community of researchers; a development which has had positive effects on applied economic research. Ultimately though, economic policy will only benefit from these improvements to the extent that quantitative techniques are applied in an appropriate way and results are interpreted carefully.

This thesis combines several of the above mentioned aspects. It uses experimental methods to shed light on fundamental aspects of human behaviour and discusses the use of econometric techniques in the policy relevant areas of development economics and international finance.

The thesis consists of two parts. The first part is an experimental analysis of a repeated 'trust game', and is organized in three chapters. It investigates the motives that determine the evolution of trust in economic interactions and discovers significant differences in attitudes towards trust according to gender and cultural background. The two self-contained chapters in the second part look into the performance of econometric techniques, and discuss their application to schooling decisions in developing countries and the properties of exchange rates. The remainder of this introduction gives a non-technical summary of the thesis and its main contributions.

In the original form of the trust game, one player, the sender, is given a certain amount of money. He can then decide how much he wants to transfer to another player, the so-called receiver. On the way to the receiver's account, the amount sent is tripled. Then it is the receiver's turn to decide how much of the tripled amount he wants to transfer back to the sender, without any obligation to do so. The money sent can be interpreted as an investment in a project, the increase as the return on investment. The project itself is managed by the receiver who decides how to divide the surplus. Making a transfer is an action involving trust because the sender deliberately increases his vulnerability to the receiver's action. A high return by the receiver is interpreted as a reward for the trust that the sender put into the receiver and is an indication of the receiver's trustworthiness.

The trust game is particularly interesting because the predictions of game theory are in sharp contrast to the observed outcome in the experimental laboratory. The prediction says that a receiver will not return any money, because every cent returned reduces his own payoff. Consequently, anticipating this reaction, a rational sender would never transfer anything in the first place. However, experiments have shown that players trust their anonymous counterparts blindly and receivers in turn reward senders for this behaviour even if they will never actually meet their opponent.

The setup of the experiment analyzed in this thesis modifies the basic setting in three ways. First, players are matched in groups of five. Every sender *can choose* one receiver out of the four other players in his group. The players' nationality, gender, age and the number of siblings are public information within the group. The second modification is that all players assume the roles of *sender and receiver* because they choose each other simultaneously. The third alteration is that the same group *interacts repeatedly* for six times.

Hence, a dynamic environment is created in which trust can emerge as the result of a repeated interaction. These modifications, however, do not alter the predictions of standard game theory; repeatedly applying the argument sketched above leads to the same result that trusting an opponent is not rational.

Chapter 1 looks at the general behaviour of subjects in this experiment.[1] Participants show a high degree of trust in that they transfer and return substantial amounts throughout the game. The chapter provides evidence for alternative motives that might be causing this non-rational behaviour. It finds that players behave reciprocally by rewarding good behaviour, such as someone choosing them as a player and making high transfers, with high returns and future contacts. The chapter also shows how players learn from previous playing experience. For example, if players see that a certain action leads to a high payoff, they tend to repeat it and they do the opposite if they learn that their behaviour was not successful. To a certain degree players also behave rationally because they transfer less towards the end of the game. In conclusion, Chapter 1 finds evidence for motives that – in addition to rationality – give a more complete description of players' behaviour in the game.

The most important extension of the game analyzed here with respect to previous studies lies in the option of choice added to the trust game. In the present setting players choose their preferred partner among four players. The reasons driving the choice are the subject of the analysis in the remaining two chapters of the first part of this thesis. What makes this setup especially interesting is that not only the characteristics of the decision maker and his preferred choice are observed, but also the characteristics of the alternatives that were not chosen. Using this information, Chapter 2 finds that male players have a higher propensity to contact female players as opposed to male players, while female players do not discriminate in their choices. However, this discrimination does not pay off in terms of higher returns made by women and is therefore not rational from a payoff point of view.

Perhaps the most striking finding of the experiment is the discrimination that occurs along the lines of nationality. Dividing the fifteen European nationalities that participated in this experiment into two geographical regions (North and South) evidence is provided in Chapters 2 and 3 that southerners are discriminated against, and mainly so by northerners.[2] As shown in Chapter 3, this discrimination builds up in the course of the experiment rather than dying out with experience. The reason for this discrimination is that rather

1 This chapter is joint work with Andrea Ichino, Karl Schlag and Eyal Winter.
2 Chapter 3 is joint work with Andrea Ichino, Karl Schlag and Eyal Winter.

than for not being trustworthy (*i.e.* having a low propensity to reciprocate to senders by making a generous payback for a transfer received), southerners are being punished for their own low level of trust, meaning that they have a low propensity to contact another player with a generous transfer in the first place. For this reason southerners end up leaving the game with lower payoffs.

The conclusions that can be drawn from this experiment are twofold. First, observed behaviour differs substantially from predictions of standard economic theory. Individual behaviour is guided by motives such as reciprocity, and individuals adapt their behaviour as a response to previous experience. The existence of diverse motives and learning mechanisms not only points to the rather incomplete view of human behaviour in standard models, but also puts into question predictions of models that rely exclusively on individual utility maximization.

Second, the analysis shows that individual characteristics such as cultural background and gender influence attitudes towards trust, and that these variations lead to significant differences in the way individuals treat each other. While the aim of the analysis is not to give a conclusive answer to the question of why such discrimination occurs, its mere existence has wide reaching implications, especially in the European context. Regional differences in Europe regarding standards of trust – as established in the experiment – may pose impediments to a process of European unification that is characterized by political and economic uniformity. In more practical terms, the results point at potential difficulties encountered by firms when operating in a culturally diverse Europe, as trust is an important factor in business relationships.

The second part of the thesis combines methodological issues with concrete economic problems, and each chapter has important policy implications aside from making a methodological contribution.

Chapter 4 deals with the evaluation of a large scale poverty reduction programme undertaken by the Mexican government in the late nineties, called PROGRESA. Low secondary school enrolment ratios are recognized as a major obstacle to breaking the poverty cycle. No education leads to few employment opportunities, which in turn implies a low income. Children in poor countries are often deprived of school education and obliged to contribute to household income through work. To increase secondary school enrolment, the Mexican government offered a cash payment to poor households that was paid *conditional* on the school attendance of the children. The rationale is that upon receiving income support families can afford to send their children to school.

What makes PROGRESA an exceptional case is that this programme was implemented as a randomized experiment. This means that part of the targeted population was not offered participation in the programme. For this reason, the effect of the programme can easily be identified simply by comparing its effect on those that received the transfer with what happened to those that did not. Social experiments such as PROGRESA are rare because policy makers, for ethical and social reasons, are often reluctant to conduct such field experiments. But in most cases the underlying economic problem – such as schooling decisions in the Mexican case – is far too complex to be addressed in an experimental setup in a campus laboratory. However, the need for evaluating such programmes remains, and it is difficult to identify the effect of policies without a proper control group. Microsimulation methods are one way to overcome the lack of experimental data. Starting from a single individual as the decision making unit, microsimulation methods simulate how individuals would react if they faced certain situations – thus making an experimental setup unnecessary. Chapter 4 combines a microsimulation method with the social experiment PROGRESA. In particular, a microsimulation of the programme is carried out and then compared to the real effect of the policy. This procedure offers a benchmark for the simulation technique and enables discussion of the accuracy of the microsimulation technique and hence an assessment of its usefulness for policy advice.

The analysis shows that the prediction of the microsimulation method comes close to the real effect. The overall performance of the technique is satisfactory and some critical issues are discussed in the chapter. But in addition to validating a microsimulation technique, the analysis in Chapter 4 gives empirical evidence for the functioning of conditional cash transfer schemes in practice. The data support the view that lack of education is often the result of financial constraints faced by poor households in developing countries. The demand for schooling is high, but only if households get income support can they afford to send their children to school.

The chapters dealt with so far look at the immediate consequences of individual behaviour. But economic theory is also concerned with the relation of economic variables at the aggregate level. For example, the purchasing power parity theory states that the same basket of goods should cost the same in any country, once differences in price levels and currencies are accounted for. The rationale for this theory is that if any deviations from the parity were persistent, arbitrage opportunities would emerge and as a consequence of trade the relative prices of the goods would eventually move back to equilibrium. This implies that observed exchange rates, corrected for price level differences, should fluctuate around a long-run equilibrium value.

But observable data is always influenced by random shocks, making it difficult to infer from the exchange rate of one country a 'typical' rate, and thus to find evidence for or against the purchasing power parity theory. Investigating the exchange rate of many countries simultaneously is one way of broadening the basis for the argument, and data sets with a cross section and a time series dimension are increasingly being used to investigate macroeconomic phenomena. But the use of both the cross section dimension *and* the time dimension leads to certain statistical particularities, which are the subject of the analysis in Chapter 5. As is the case for exchange rates, data may not always be independent across countries; exchange rates of different countries are exposed to the same shocks, creating sectional dependence in the data. This dependency causes problems for several panel estimators and its effects are analyzed in Chapter 5 using simulation techniques.

In particular, the chapter investigates and demonstrates the distortions that testing procedures may have if applied to macroeconomic data sets that do not fulfil the assumption of cross sectional independence. The chapter highlights the importance of carefully checking the appropriateness of econometric techniques prior to their implementation and discusses how this can be done in the case of the tests considered.

Part I

An Experimental Analysis of the Repeated Trust Game

CHAPTER 1

Learning to Trust: Analyzing Motives in a Repeated Trust Game[1]

1.1 Introduction

The purpose of this paper is to investigate trust and trustworthiness in a dynamic setting. We set up an experiment where trust can emerge as the result of an experimentally controlled repeated interaction between individuals. Hence, we do not only study the general propensity of people to trust, but also the motives that determine the evolution of trust.

In the trust game a player is given 100 units of the experimental currency and is allowed to send some of it to a different player. During the transaction the transferred money is tripled. Finally, the recipient is allowed to return part of the tripled transfer without any obligation to do so. The money sent can be interpreted as an investment in a project, the increase during the transfer as the return on investment. The project is managed by the recipient who decides how to divide the surplus.

The way that game theory analyzes this trust game is to invoke backwards induction. For any given amount transferred, the receiver is best off not returning anything. Knowing this, the sender will not send anything in the first place. The outcome of this behaviour is inefficient, and is reminiscent of the Prisoners' Dilemma where similarly an inefficient outcome is predicted by game theory. Any efficient outcome (equivalent here to maximizing the sum of the payoff of the sender and of the receiver) is characterized by the

1 This chapter is joint work with Andrea Ichino, Karl Schlag and Eyal Winter.

sender transferring all 100 units (so the recipient receives 300 units). In our experiment we implement a repeated trust game where players have the opportunity to select a new opponent in each round. However, given that there is a finite number of rounds (6 in our experimental design), the backwards induction argument still yields the same result. No player should ever send anything.

The rational prediction is mainly a theoretical benchmark, and experiments show that subjects trust (send money) even when the trust game is played only once. Berg et al. (1995) find that subjects send slightly above 50 points and return slightly less to the sender and keep over 100 points for themselves. Among their subjects it was not rational, given the behaviour of the recipients who on average return 47 points, to transfer anything. Burks et al. (2003) show that if two subjects get money to send to each other simultaneously (so both are sender and receiver) then subjects send again about 50 points but return much less, namely on average 24 points. Here it is even less rational to send money, because subjects seem even less trustworthy when they are both sender and receiver. Our design is related to Burks et al. (2003) because all subjects are senders and possibly also receivers. It is also related to Cochard et al. (2000) as we repeat the game a finite number of times.

As rationality is a poor predictor we test for other motives such as reinforcement, reciprocity and directional learning. We find that much of the observed behaviour in the game can be explained by the two motives reciprocity and reinforcement learning. Players reward opponents for their choices and their actions if their behaviour was favorable. This is visible in the choice, the transfer made and the ratio returned. In addition, payoff oriented reinforcement of actions is also observable. Players are more likely to repeat their actions if they have proven successful. Finally, the end game effect that can be observed both in the transfers made as well as in the ratio returned is indicative for some degree of rationality.

The reminder of the paper is organized as follows. Section 1.2 relates our experiment and the main findings to the existing literature. The experimental design is briefly described in Section 1.3. Section 1.4 presents some general descriptive statistics on the game. Different behavioural motives are briefly discussed in Section 1.5 before an econometric analysis is undertaken in Section 1.6. The last section concludes and discusses directions for further research.

1.2 Related Literature

While sociologists mainly use attitudinal surveys on rather vaguely defined concepts of trust and trustworthiness, economists have recently been trying to be more precise on the issue and its conditioning factors. Glaeser et al. (2000) combine survey data and experimental data in an attempt to quantify the general perception of trust towards different groups surrounding an individual.

There is evidence that trust and trustworthiness are related to the sociological background of people. For example, Buchan et al. (2000) find some support for the relationship between trust and social distance across countries. Fershtman and Gneezy (2001) find different levels of trust according to the opponents' origin. Croson and Buchan (1999), among others, identify gender as yet another determinant for trust, with women being trusted more than men. These results stress the importance of controlling for confounding factors when the emergence of trust in an economic interaction is analyzed.

Our experimental design combines several elements of previous studies. The basic trust game with one sender and one randomly matched receiver is known from the study by Berg et al. (1995). In this study pairs are matched with assigned roles as sender and receivers to play a one shot trust game. We follow the extension by Burks et al. (2003) that both players assume the role of a sender and receiver at the same time. However, contrary to this study, this was known to the players from the outset. We also combine the element of a repeated game as analyzed by Cochard et al. (2000), but also run a control treatment with one shot interactions. In addition, we add the element of a free choice, which to our knowledge has not been investigated in this context. Our results compare nicely to the existing literature as indicated by Table 1.1.

With experimental economics being a rather new field in economics, a thorough econometric analysis of experimental data is more the exception than the rule. Numerous studies content themselves with basic descriptive statistics. The advantages of an econometric analysis is that confounding factors can be controlled for, preventing the premature interpretation of results. However, to link the experimental setup to the correct econometric specification is not always an easy task. For example, with the exception of very simple games, the derivation of a likelihood function is intractable for more complicated settings. Hence, the correspondence between the theoretical model and the empirical specification is not always perfect. An exception in this context is the analysis by El-Gamal and Grether (1995) who are able to translate their (simple) game one-to-one into a likelihood function, estimate and identify the relevant parameters. Identification is a particular problem in the context of behavioural economics. As Manski (2002) points out, several behavioural hy-

Table 1.1: Results and related literature

	N	*avg. sent (0-100)*	*average returned*	*main features*
Berg et al. (1995)	32·2	52	47[a]	assigned roles, one shot
Burks et al. (2003)	22·2	65	85[a]	assigned roles, one shot
Burks et al. (2003)	20·2	47	24[a]	S and R, one shot
Cochard et al. (2000)	30·2	42	39%[b]	assigned roles, one shot
Cochard et al. (2000)	16·2	75	56%[b]	assigned roles, repeated
this study				
repeated	110	76	54%[b]	S and R, choice,
random assignment	110	67	38%[b]	random assignment

Note: Assigned roles means subjects were assigned roles as sender or receiver. S and R means subjects act both as a sender and receiver, N is the total number of participants, indicating the number of pairs that played. [a] amount sent back, [b] ratio returned, conditional on having received a positive amount.

potheses might be observationally equivalent, making it impossible for the econometrician to distinguish between them. Our aim is to characterize typical behaviour at different stages of the game. We confine ourselves to find empirical support for or evidence against such hypotheses controlling for confounding factors.

1.3 Experimental Design

The reader is referred to Appendix A.1 for a full documentation of the experimental setup, including a transcript of the instructions and screenshots. Here we describe only the main features. A total of 110 subjects participated in a computerized setup in three sessions.[2] Each of the sessions consisted of six treatments. In each treatment, subjects were randomly matched in groups of five players to play the repeated trust game described below. Each player had information about the nationality, the gender, the age and the number of siblings of the four opponents in the same group.

Free Choice Treatments

Treatments one to four and treatment six were so called 'free choice' treatments (f1-f5). In stage one of the game the subject decided *who* and *how much* of his initial endowment of 100 to transfer to a chosen player. Not making any transfer was also an option. In stage two, subjects saw *who* of the

2 Using the Z-Tree software (Fischbacher, 1999).

other players had chosen them and *how much* each of them had transferred. In addition, this amount was shown multiplied by three. Subjects could then decide how much to transfer back to each player they had received a transfer from. In stage three subjects were presented a summary of all transfers and returns they had been involved with that happened in this period.[3] The three stages were repeated six times. Then, groups were reshuffled and a new treatment was played.

Control Treatment

Between the fourth and the fifth free choice treatment subjects were informed via the screen about a small change in the game. They were again matched in groups of five players, but instead of being able to choose a fellow player, they were *randomly assigned* one of the fellow players (see Figure A.5 in the Appendix). We also call this the predetermined treatment. The random assignment was implemented by selecting one of the choices with equal probability. Hence, it was still possible that the same player receives transfers from various players or from no player, but these events were random. In every period of this treatment players faced a new, random choice of the same group. After this treatment, subjects played a last free choice treatment.

1.4 Descriptive Statistics

The following statistics are organized around the course of the game, starting with statistics regarding the choice, then the amount transferred, and lastly the amount returned. They provide a rough description of the playing behaviour. Empirical evidence and interpretation of types will be discussed in Section 1.6. Unless indicated differently, the statistics do not include the control treatment.

Choice

In each period subjects had the option to choose one of the four players in their group to transfer points to. This group of players remained unchanged for six consecutive periods. Table 1.2 summarizes by treatment and period how often subjects decided not to change their chosen partner. The analysis, period by period, shows a slight increase in periods 2-5 from 53 percent to 57 percent of the players who stay with the same partner, but this share drops to just 47 percent in the last period.

3 See screenshots in Figures A.2, A.3 and A.4 in the Appendix.

Table 1.2: Persistence of choice of player 1

	period of treatment					
treatment	2	3	4	5	6	*total*
f1	0.42	0.47	0.50	0.55	0.47	0.48
f2	0.51	0.55	0.58	0.54	0.46	0.53
f3	0.47	0.52	0.53	0.56	0.55	0.53
f4	0.51	0.58	0.57	0.64	0.45	0.55
f5	0.52	0.53	0.61	0.58	0.41	0.53
total	0.49	0.53	0.56	0.57	0.47	0.52

Note: The table shows for each treatment/period combination the fraction of players that chose the same partner as in the previous period. Each treatment/period combination was played 110 times.

There was also considerable persistence in the choice of partner exceeding one period. From Table 1.3 it can be seen that while the in the majority of cases a player was chosen only once (58 percent), 24 percent of the players remained with the same choice for at least two periods or more. 2 percent did not change the player throughout an entire treatment. Conditional on switching to a new player and having received a positive transfer in the previous period, 62 percent chose a player from which they had received before.

Table 1.3: Persistence of choice of player 2

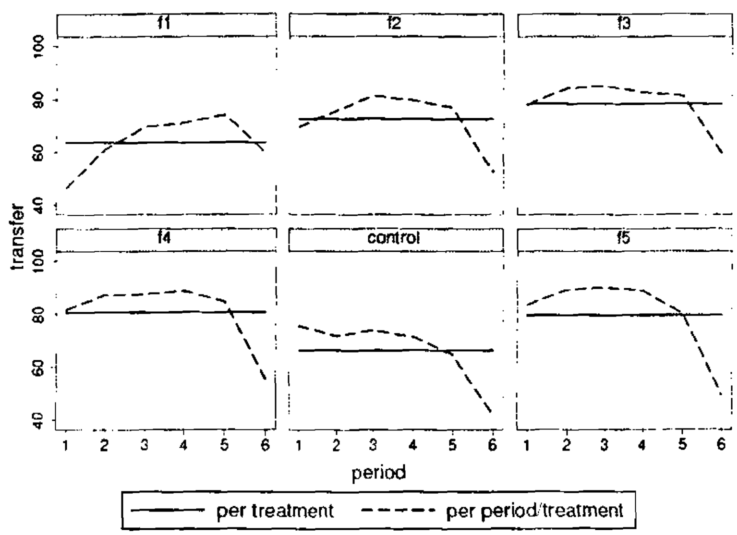
	number of consecutive periods					
	0	1	2	3	4	5
absolute	1899	603	339	225	156	78
fraction	58	18	10	7	5	2

Note: The table shows the absolute number and the fraction of cases in which a player chose another player for the number of consecutive periods displayed in the first row. This table includes observations for all periods.

Transfer

The average transfer was 76 tokens. Figure 1.1 shows how the transfer depends on period and treatment, where the control treatment is also included. It reveals a significant drop in the last period of each treatment. All free choice treatments exhibit the same effect over time. On average, the amount transferred increases from 72 in period 1 to 83 in period 4, decreases slightly to 80 in period 5 before it drops to 56 in the last period, well below the value of the starting period. The average amount transferred rises from 64 in the f1 treatment to 81 in f5 treatment (solid line in Figure 1.1). The average in the control treatment is 66 and therefore as low as the first treatment. The control treatment was the fifth treatment, and as can be seen from Figure 1.1 the transfer behaviour in the sixth and last treatment resumes the pattern of the previous free choice treatment. This shows that subjects understood well the difference between the control and the free choice treatments.[4]

Figure 1.1: Average transfer per period and treatment



There is a clustering of transfers at certain values, as Table 1.4 illustrates. In 56 percent of the cases the full amount of 100 points was transferred. A second point mass is at the value of 50, which was the amount transferred in 9 percent of the cases in the free choice treatment.

Return Ratio

The ratio a player got back from his initial transfer is defined as $r = G/(3 \cdot t)$, where G is the amount returned and t is the initial transfer which is multiplied by three upon arrival on the opponent's account. Hence, $r \in [0, 1]$. As can be seen in Figure 1.2, the average return ratio (0.54) does not have such a large variation between the free choice treatments (0.51-0.59) but is significantly lower in the control treatment (0.39). The end game effect is also quite visible in the free choice treatments, where the ratio drops from an average of 0.58

4 For a histogram of transfers period by period, see Figure A.6 in the appendix.

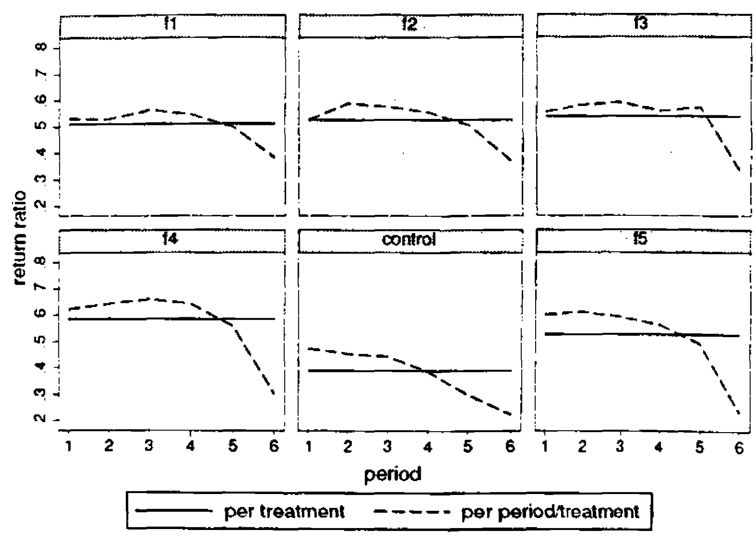
Table 1.4: Distribution of transfer

	percentage	
$t =$	*free choice*	*predetermined*
0	7	18
$1 \leq t \leq 49$	12	7
50	9	12
$51 \leq t \leq$	17	17
100	56	46

Note: The table reports the percentage share of transfers t that fall in the range indicated in the first column.

in periods 1-5 to 0.33 in the last period. In the predetermined treatment the return ratio around is around 0.45 for the first three periods and then drops steadily to just 0.22 in the last period.

Figure 1.2: Average ratio returned per period and treatment



The return ratio also clusters at certain values, as shown in Table 1.5. In the free choice treatments the biggest point masses are at 1 and at 2/3, followed by 1/2 and 0. Looking at Figure A.7 in the appendix we see that the end game effect is driven by the large share of zero return ratios (37 as compared to 8 percent in periods 1-5). In comparison, we find 48 percent of return ratios equal to zero in the last period of the predetermined treatment.

Both the transfer sent and the return ratio received increased with number of consecutive times a sender contacted a receiver, as reported in Table 1.6.

Table 1.5: Distribution of the return ratio

	percentage	
$r =$	free choice	predetermined
0	12	20
$0 < r < 1/3$	6	9
1/3	9	13
$1/3 < r < 1/2$	7	10
1/2	16	22
$1/2 < r < 1/2$	7	7
2/3	18	13
$2/3 < r < 1$	7	3
1	18	3

Note: The table reports the percentage share of return ratios r that fall in the range indicated in the first column.

Average transfers increase substantially from 68 to 90, and the return ratio increases from 0.51 to 0.62 after 3 consecutive choices. Interestingly, if a player was chosen for four consecutive times or more, the return ratio drops again.

Table 1.6: Transfer and return ratio for consecutive choices

	number of consecutive periods				
	0	1	2	3	4+
t	68	81	88	90	90
r	0.51	0.60	0.60	0.62	0.53

Note: The table reports the average transfer t and the average ratio returned r for the number of consecutive periods that a sender chose the same receiver.

Payoff

The payoff π is defined as the payoff a sender gets from a single interaction, i.e. $\pi = 100 - t + G$. From Table 1.7 we see that the average payoffs are higher in the free choice treatments.

This section can be summarized as follows. Players do not choose their counterparts at random, instead, they show a substantial reluctance to switch to new players. If so, they seem to prefer those players who they have been chosen by before. Previous interaction seems to have a positive effect both

Table 1.7: Payoff in different treatments

	period of treatment						
	1	2	3	4	5	6	*total*
free choice	155	168	171	166	150	101	152
predetermined	127	123	125	112	95	85	111

Note: The table displays the average payoffs $\pi = 100 - t + G$ in each period for the free choice and the predetermined treatment separately.

on the transfer and on the ratio that is returned to a sender. One can further see from the analysis that players transfer and return more in each repetition of the game up to period 4. In period 5 the end game effect starts, which is visible by stagnating or slightly decreasing transfers. The end game effect is strongest in the last period where substantially less is transferred and returned.

1.5 Motives and Behavioural Theories for Making Predictions

Several motives and learning theories compete in explaining the way people behave. Below we present some of the most common, namely reinforcement learning, reciprocity, directional learning and rationality. We will then develop hypotheses based on these motives for choice, transfer and return and test them. Our approach is to utilize the qualitative predictions of these motives rather than fitting explicit functional forms derived from them.

Reinforcement Learning (RIF)

Reinforcement Learning describes success oriented behaviour according to which the subjects choose actions or strategies depending on how successful they were in the past. Success is measured in terms of earned payoffs. Originating from psychology and biology, where it has been widely studied in both humans and animals, this learning strategy has recently been introduced to economics (Erev and Roth, 1998). Accordingly, individuals treat the environment as a decision, do not utilize information on how payoffs are generated and in particular ignore the fact that their opponents are also making choices. The subject is assumed to randomize over its actions according to some distribution. Positive reinforcement means that when facing the same decision again, the same action is chosen with a higher probability. Typically, positive reinforcement is more likely the more successful an action was. Negative reinforcement in turns means that the same action will be chosen with a smaller probability.

Reciprocity (RCP)

Reciprocity is a motive oriented behaviour. Cooperative and friendly behaviour is rewarded and unfriendly or non-cooperative behaviour is punished, possibly at a cost. Falk and Fischbacher (1999) provide a formal definition of reciprocity in a specific game theoretic setting. It is important to highlight that altruism, in contrast to reciprocity, is an unconditional attitude (see e.g. Cox (2002) and Falk (2003)), whereas reciprocity conditions on the actions of others.

Directional Learning Type (DLT)

The Directional Learning approach was developed by Selten and Stoecker (1986) for simultaneous move games. According to this theory, after some time people evaluate their experience and adjust their behaviour according to what would have been a better decision provided that the opponents would not change their behaviour. DLT does not make any predictions about the quantitive change of behaviour, but indicates the qualitative direction of the change.

Rationality (RTN)

The finitely repeated trust game has a unique subgame perfect Nash equilibrium in which each player sends zero in each period and returns zero whenever something positive is received. This can be derived using the standard procedure of backwards induction, whereby anticipating a zero return in the last period from a rational player, no player will ever transfer any points in the preceding period, and so forth.

1.6 Econometric Analysis

1.6.1 The Choice of a Player

The motives outlined in Section 1.5 (RIF, RCP, DLT, RTN) will be used to predict how choice probabilities should relate to specific histories. Hypotheses are formulated from the perspective of a sending player.

Hypothesis RCP 1 *One is more likely to choose a player from which a transfer was received in the previous period and who transferred a lot.*

Hypothesis RCP 2 *One is more likely to choose the same again if that player returned a lot.*

Hypothesis RIF 1 *One is more likely to choose the same again if the payoff was high.*

Hypothesis DLT 1 *One is more likely to choose the same again if that player returned more than one sent (returned ratio is greater or equal than 1/3).*

The framework in which the theoretical predictions will be addressed is the conditional logit model (McFadden, 1973).[5] The model is motivated using a random utility model representation. Define

$$U_{ijt} \text{ as the utility of } i \text{ if } i \text{ chooses } j \text{ in time period } t \text{, and}$$

$$d_{ijt} = \begin{cases} 1 & \text{if } i \text{ chooses } j \text{ in } t, \\ 0 & \text{otherwise.} \end{cases}$$

The time index t stands for the six periods of the game. Conditional on participating in the game (*i.e.* not making a zero transfer), each player has four choices, $j = \{1, 2, 3, 4\}$, in each period. The four choices are mutually exclusive and exhaustive. The basic random utility model is defined as:

$$U_{ijt} = \alpha d_{ijt-1} + \delta d_{jit-1} + \lambda p_{ijt-1} + \nu X_{ij} + \epsilon_{ijt} \tag{1.1}$$

for $j = \{1, 2, 3, 4\}$. Following the notation above, d_{ijt-1} means that player i has chosen j in the previous period. Similarly, d_{jit-1} means that player j has chosen player i in the previous round. Finally, p_{ijt-1} means that i and j have formed a pair in the previous period and is the interaction of the previous two variables. Note that, put together, these variables cover all possible cases in which there was interaction, as compared to the case in which the players have not interacted in $t-1$. The other covariates X_{ij} include the remaining choice specific characteristics such as gender, nationality[6] (both interacted with the corresponding attributes of i), age, and siblings. Notice that the

5 For a discussion of the conditional logit model in this context see also Section 2.2.1 in Chapter 2.

6 Throughout the paper we group the nationalities into participants from North and participants from South. Further analysis of the effect of nationality on the playing behaviour can be found in Chapter 3.

previous choice of i is interpreted as a characteristic of the choice j in t. By the same token, the fact that a player was chosen by some other player in period $t-1$ becomes a characteristic of that player in t. Hence, previous playing behaviour can be seen as observable choice specific attributes in t.

So far the model only accounts for the choice relating variables, e.g. if a a player was chosen or not. In an additional set of estimates, the random utility model presented in equation (1.1) will be enriched by adding variables characterizing the previous behaviour in more detail. To this end, the choice variables defined above will be interacted with variables indicating a specific behaviour, as outlined in the hypotheses. Define t_{ijt-1} as the transfer t from i to j in period $t-1$ and G_{ijt-1} as the amount player i got back from player j in period $t-1$. Then,

$$r_{ijt-1} = \frac{G_{ijt-1}}{3 \cdot t_{ijt-1}} \text{ is the ratio } i \text{ got back from } j \text{ in } t-1$$

$$\pi_{ijt-1} = 100 - t_{ijt-1} + G_{ijt-1} \text{ is the payoff of player } i \text{ in } t-1 \ .$$

In addition, the variable t_{jit-1}, the amount i received from j in $t-1$ will be used. Notice that these variables only take positive values if the respective choice specific dummy variables defined above take the value one and are zero otherwise. For example, t_{jit-1} only takes positive values if d_{jit-1} is one.

Player i chooses player j if this yields highest utility. Hence,

$$\mathrm{P}(d_{ijt} = 1) = \mathrm{P}(U_{ijt} > U_{ikt}) \, \forall k \neq j.$$

Table 3.10 and 1.9 contain the estimation results of various specifications of the conditional logit model. For convenience, the choice variables are represented using arrows, where d_{ijt-1} is represented by a dashed arrow $i \dashrightarrow j$ and the pair variable by a double arrow $i \leftrightarrow j$.

Consider specification C1. This model disregards any success or failure of previous choices and forms the basis for the following analysis. It is evident that having chosen a player before and having been chosen by a player makes it more likely to choose that player again. The effects are of the same order of magnitude, with the effect of $i \dashleftarrow j$ being slightly bigger. Interestingly, the effect of having formed a pair does not significantly alter the choice probabilities, suggesting that there is no pair-specific effect. In the following specifications this variable is dropped. Notice that this simple model can predict 59 percent of the choices correctly.

Table 1.8: Choice: conditional logit estimation results 1

	C1	C2	C3	C4
$i \leftrightarrow j$	.19	.	.	.
	(.13)			
$i \dashrightarrow j$	1.16	1.18	-.46	-.61
	(.06)***	(.04)***	(.11)***	(.12)***
$i \dashleftarrow j$	1.37	-.05	.23	.21
	(.07)***	(.17)	(.17)	(.17)
$i \dashleftarrow j \cdot t_{ji}$	.	.02	.02	.02
		(.002)***	(.002)***	(.002)***
$i \dashrightarrow j \cdot r_{ij}$	.	.	2.93	.
			(.19)***	
$i \dashrightarrow j \cdot \pi_{ij}$	.	.	.	.01
				(.0007)***
Obs.	10160	10160	10160	10160
Pseudo R^2	.25	.26	.30	.31
Correct Predictions	.59	.59	.62	.62

Note: All variables refer to the previous period. $i \dashrightarrow j$ means that i has chosen j, and $i \dashleftarrow j$ means j has chosen i. The interaction of both is denoted by $i \leftrightarrow j$. t_{ji} is the transfer received from j, r_{ij} the return ratio and π_{ij} the payoff. Reported values are coefficients, standard errors in parenthesis. *, ** , *** denote significance to the 10, 5 and 1 percent level of a test that the coefficient is zero. Control variables included are: age and siblings of all 4 players, gender and nationality of i interacted with the attributes of j. Pseudo R^2 is the percent of variance explained by the model compared to a model which includes a constant only. Correct predictions indicates the share of observations in which the highest estimated probability $\hat{p}_j$ coincides with the actual choice.

Hypothesis RCP 1, which says that the probability of choosing a player is increasing in the amount received from that player is addressed in specification C2 in the table, where the variable which indicates that $i \dashleftarrow j$ is multiplied by the amount transferred. Indeed, the likelihood of choosing a player who transferred previously is increasing in the transfer received. Hence, hypothesis RCP 1 finds empirical support.

Hypothesis RCP 2, which states that choosing the same player again is more likely if that player returned a lot, is addressed in specification C3, where the ratio the player returned in the previous period is added. Notice that this specification also controls for the amount received by a player in the previous playing round. The significance of the interacted variable $i \dashrightarrow j \cdot r_{ij}$ gives empirical support for Hypothesis RCP 2. The inclusion of the return ratio increases the predictive power of the model to 62 percent. Notice, however, the significant negative coefficient of the indicator variable $i \dashrightarrow j$. This means that only if the return ratio is bigger than 0.17 players will reciprocate behaviour and choose that player more likely.[7]

In specification C4 we then add the payoff resulting from this interaction. Here we find that reinforcement depends on the success of the previous action: if the payoff was above 60, we find evidence for positive reinforcement as the probability of choosing the same again is higher, confirming RIF 1.[8] Negative reinforcement is triggered when the payoff is below 60, which is the case in 13 percent of the cases. We do not include both the returned ratio and the payoff in the same regression because $i \dashrightarrow j \cdot r_{ij}$ and $i \dashrightarrow j \cdot \pi_{ij}$ are highly collinear.

Table 1.9: Choice: conditional logit estimation results 2

	C5	C6
$i \dashleftarrow j$	.26 (.17)	.14 (.17)
$i \dashleftarrow j \cdot t_{ji}$	.02 (.002)***	.02 (.002)***
$\mathbf{1}\{i \dashrightarrow j \wedge r_{ij} \geq 1/2\}$	1.63 (.05)***	.
$\mathbf{1}\{i \dashrightarrow j \wedge r_{ij} < 1/2\}$	.12 (.09)	.
$\mathbf{1}\{i \dashrightarrow j \wedge r_{ij} \geq 1/3\}$	.	1.38 (.05)***
$\mathbf{1}\{i \dashrightarrow j \wedge r_{ij} < 1/3\}$	.	-.28 (.15)**
Obs.	10160	10160
Pseudo R^2	.29	.28
Correct Predictions	.61	.61

Note: See notes to Table 3.10. The symbol "$\wedge$" is the logical "and" operator and $\mathbf{1}\{...\}$ is the indicator function which takes value one if the expression inside the parenthesis is true. For example, $\mathbf{1}\{i \dashrightarrow j \wedge r_{ij} \geq 1/2\}$ takes value one if i has chosen j and in addition the ratio returned was greater or equal than 1/2.

Hypothesis RCP 2 is a more general version of DLT 1. While the former just makes a general statement about how the return ratio affects choice, the latter is very specific in determining a break point at 1/3. To shed more light on the functional form, specifications C5 and C6 in Table 1.9 test for a

7 To find the point at which the return ratio contributes positively to the probability of choosing a player (assuming a linear relationship) solve $-0.46 + 2.93r_{ij} \geq 0$ for r_{ij}.

8 Again, to find the point at which the payoff contributes positively to the probability of choosing a player solve $-0.61 + 0.01\pi_{ij} \geq 0$ for π_{ij}.

change in slope at 1/2 and at 1/3, respectively. A return ratio of 1/2 is the median return ratio. From specification C6 it becomes evident that players chose the same player less likely if the ratio returned previously was lower than 1/3. This can interpreted as evidence for DLT 1. For a breakpoint at 1/2 no such evidence can be found, players' probability of choosing the same player remains unaffected by the variable (see C5).

Table 1.10 summarizes the main results of this section.

Table 1.10: Choice: summary of findings

	...probability of i choosing j is higher	*evidence*
RCP 1	if i received a lot from j	yes
RCP 2	if j returned a lot previously	yes
DLT 1	if j returned more than i sent	yes
RIF 1	if i had a high payoff	yes

1.6.2 The Amount Transferred

After a player has chosen who to send to, players could choose how much of their endowment of each period to transfer. Our motives will be used to predict changes in the amount transferred. These motives are formulated analogously to the choice setting, where an increase in transfers is the equivalent action to increasing the probability of choice.

Hypothesis RCP 3 *Conditional on choosing a player from which a transfer was received in the previous period, the current transfer is increasing in that transfer received.*

Hypothesis RCP 4 *Conditional on choosing the same player again, the transfer is increasing in the ratio returned in the previous period.*

Hypothesis RIF 2 *Conditional on choosing the same player again, the transfer is increasing in the payoff received in the previous period.*

Hypothesis DLT 2 *Conditional on choosing the same player again, the transfer is higher (lower) if the ratio returned in the previous period is greater (smaller) than 1/3.*

Hypothesis RTN 1 *The transfer is lower in the last period.*

The following analysis gives econometric evidence for the hypotheses outlined above, controlling for confounding factors. The test idea is to see if the events, which are characteristic for a certain motive, lead to a significant difference in transfers. The framework in which the hypotheses will be addressed is

$$t_{ijt} = \alpha d_{ijt-1} + \beta d_{jit-1} + \delta X_{jt} + \eta Z_{it} + u_{ijjt},$$

which forms the basis for the analysis. t_{ijt} is the transfer sent from i to j in period t. The variable d_{ijt-1} and d_{jit-1} are defined as in the previous section. The matrices X_{jt} and Z_{it} contain a set of j and i specific characteristics, respectively, and u_{it} is a random error component. According to the hypotheses, the choice variables d_{ijt-1} and d_{jit-1} will be interacted with variables that characterize previous playing behaviour such as the amount transferred or the ratio returned. Because there are repeated observations for the same individual in the sample, the standard errors are corrected for within-individual correlations.[9]

To facilitate the reading of the tables, the following notation is introduced. Let S denote the sender and R the receiver.[10] Consistent with the notation above,

- $S \dashrightarrow R$ denotes that S has chosen R in $t-1$
- $S \dashleftarrow R$ denotes that R has chosen S in $t-1$
- $S - R$ denotes that S and R had no interaction in $t-1$.

Notice that these variables sum up to one. In the subsequent analysis, $S - R$ will be the omitted variable. The hypotheses distinguish between the behaviour towards a player according to whether he is the same or different choice as in the last period. Hence, the variables are interacted with a variable that indicates whether the same choice was made in t and $t-1$. If a variable corresponds to the set of same choice or the set of different choice will be indicated by s and d, respectively.

Consider Table 1.11. Specification T1 forms the starting point for the analysis. Transfers are higher if the same subject is chosen again, on average by 7 points. Having been chosen by a player previously increases transfers sent to that player on average by 11 points. In specification T2 the amount received is interacted with the indicator that S was chosen by R ($S \dashleftarrow R$). In order

9 See Moulton (1986).

10 The notation differs from the (i, j) notation used in the analysis of choice, because at this stage of the game players have already chosen a particular receiver among the j possible options.

Table 1.11: Transfer: estimation results 1

	T1	T2	T3	T4
$S \dashrightarrow R$	7.22 (1.54)***	6.75 (1.52)***	-.44 (3.35)	-5.1 (3.7)
$S \dashleftarrow R$	10.98 (1.53)***	-19.26 (4.56)***	-15.42 (4.41)***	-12.73 (4.36)***
$(S \dashleftarrow R) \cdot t_{ji}$	.	.33 (.05)***	.28 (.05)***	.24 (.05)***
$(S \dashrightarrow R) \cdot r_{ij} \mid s$	.	.	16.37 (3.29)***	.
$r_{ij} \mid d$	.	.	8.14 (3.19)**	.
$(S \dashrightarrow R) \cdot \pi_{ij} \mid s$	.	.	.	.1 (.02)***
$\pi_{ij} \mid d$	.	.	.	.06 (.01)***
Obs.	2540	2540	2540	2540
R^2	.21	.22	.24	.27

Note: All variables refer to the previous period. $S \dashrightarrow R$ indicates that the sender chose the receiver and $S \dashleftarrow R$ that the receiver chose the sender. t_{ji} is the transfer received from j, r_{ij} the return ratio and π_{ij} the payoff. $\mid s$ and $\mid d$ means that the variable is interacted with same (s) or different (d) choice. Reported values are coefficients, standard errors corrected for repeated observations of the same individuals in parenthesis. *, ** , *** denote significance to the 10, 5 and 1 percent level of a test that the coefficient is zero. Control variables are: gender, age, siblings, and nationality of sender and receiver, dummies for sessions, treatments and for each period.

to determine the effect of received transfers on own transfers, one has to take into account the coefficient of the variable $(S \dashleftarrow R)$, which is negative and significant, because this is the intercept of the functional relation between the amount received and transfer made. For example, with an intercept of -19 and a slope of $1/3$, the results suggest that only if the transfer received was higher than 57 points this had a positive effect on the own transfer. This is evidence for hypothesis RCP 3.

Hypothesis RCP 4, which relates transfers to the ratio returned in the previous period, also finds empirical support, as can be seen from specification T3. This is evident from the positive coefficients of the previous choice multiplied with the return ratio received and interacted with same and different choice. In this case the coefficient of $(S \dashrightarrow R)$ becomes insignificant. However, upon having received a high return ratio players increase their transfer even if they choose a different player. Hence, the *additional* reward when choosing the same player is the difference $16.4 - 8.1 = 8.3$, which is still significant. The insignificance of $S \dashrightarrow R$ points to the fact that transfers are increased even if the return ratio is small. In other words, RCP 4 also holds in the unconditional formulation.

The results for hypothesis RIF 2 are similar, for which evidence is provided in specification T4. As in the choice setting we consider RCP and RIF separately due to collinearity of payoffs and return ratios. Players do increase their transfer after high payoffs. However, again they do so *regardless* to whether they repeat their choice or not, but do more so in case of a repeated choice $(0.1 - 0.06 = 0.04)$. Similar to the RCP hypothesis above, the insignificance of $S \dashrightarrow R$ is evidence that transfer increase even if payoffs were very small. Here we speak of positive reinforcement for all payoff values. A comparison of the explanatory power of RCP 4 and RIF 2 shows that the latter one has a better fit to the data. Hence, we will refer to evidence for reinforcement as strong.

Hypothesis DLT 2 is addressed in Table 1.12. We find evidence of directional learning as the coefficient of the variable that indicates that more was returned than sent, $1\{r_{ij} \geq 1/3\}|s$, is significant. This means that when having got back less than sent, the transfer decreases significantly by 23 points $(18 - (-5) = 23)$. Notice that this effect is smaller $(7 - (-5) = 12$ points$)$ when transferring to a new player. In T7 we investigate the alternative cutoff point at 1/2. While the effect on transfers is still significant for the same choice, it is now only 2 points.

Finally, hypothesis RTN 1 is widely confirmed by the data. Table 1.13 reports the coefficients for the period dummy variables included in the regressions for two specifications - the results for any other specification are similar. It is evident that transfers increase slightly in the third period relative to the second, but drop on average by 5 points in the last period. This is the case in both regressions T3 and T5, where once the ratio returned and once the amount received was included. Hence, the end game effect is *additional* to any decreases in transfer that could have been induced by lower receipts or return ratios in the previous period.

Table 1.14 summarizes the main findings of this section.

1.6.3 The Amount Sent Back

In the last stage of the game, subjects decided how much of any amount transferred to them (multiplied by 3) to send back to the original sender. Let G_{ijt} be the amount that the sender (who is indexed by i) gets back from

Table 1.12: Transfer: estimation results 2

	T6	T7
$S \dashrightarrow R$	-5.31 (4.54)	3.43 (2.23)
$S \dashleftarrow R$	-17.32 (4.28)***	-17.15 (4.46)***
$(S \dashleftarrow R) \cdot t_{ji}$	.31 (.05)***	.31 (.05)***
$\mathbf{1}\{r_{ij} \geq 1/3\} \mid s$	17.98 (3.93)***	.
$\mathbf{1}\{r_{ij} \geq 1/3\} \mid d$	6.53 (2.26)***	.
$\mathbf{1}\{r_{ij} \geq 1/2\} \mid s$	.	5.75 (1.76)***
$\mathbf{1}\{r_{ij} \geq 1/2\} \mid d$	.	1.37 (1.81)
Obs.	2540	2540
R^2	.24	.23

Note: See notes to Table 1.11. The function $\mathbf{1}\{...\}$ is the indicator function which takes value one if the expression inside the parenthesis is true.

Table 1.13: Transfer: estimation results for period dummies

	T3	T5
period 3	1.56 (.80)*	1.35 (.80)*
period 4	1.24 (.93)	1.07 (.92)
period 5	.018 (1.05)	-.28 (1.04)
period 6	-4.95 (1.91)**	-5.26 (1.89)***

Note: Effects to be interpreted with respect to period 2. See notes to Table 1.11.

receiver j. From a receiver's perspective, the variable G_{ijt} is the amount he pays back to the sender, and will henceforth be called P_{jit}. This variable, which is naturally bounded by 3 times the amount received, $(3 \cdot t_{jit})$, will be the measure of return used in this section.

Table 1.14: Transfer: summary of findings

	...transfers are higher if	*evidence*
RCP 3	the ratio returned was high	yes
RCP 4	the receiver sent a lot	yes
RIF 2	the payoff was high	strong
DLT 2	the receiver returned more than 1/3	yes
	...transfers are lower	
RTN 1	in the last period	yes

Due to the nature of the game, when players take the decision how much to transfer back, they know already who has chosen them in this period t. Hence, in period t receivers know if they are playing in a pair or not and by whom they have been chosen.

The hypotheses are formulated from the perspective of the player that takes action, *i.e.* the receiver.

Hypothesis RCP 5 *The amount paid back is higher if the transfer received was high.*

Hypothesis RIF 3 *The amount paid back is higher if received from a player for the second time.*

Hypothesis DLT 3 *The amount paid back is lower if received from a player for the second time.*

Hypothesis RTN 2 *The amount paid back is lower in the last period.*

The choice of players in stage one of the game leads to a particular feature of the data analyzed in this section. A player might have been chosen by 0, 1, 2, 3 or even 4 other players. Hence, at each period t there are between 0 and 4 observations for each player of an amount paid back. In total, of course, there are as many observations as for the initial transfer.

The framework to be used in this section goes along the lines of the analysis of the transfer. Consider the equation for the amount ratio returned

$$P_{jit} = \alpha d_{jit} + \beta d_{jit-1} + \gamma d_{ijt-1} + \delta X_{it} + \eta Z_{jt} + u_{jit},$$

where the variables are defined as above. The variable d_{jit} indicates if in addition to having been chosen by sender i, receiver j also chose i in period t. For player j to make a move in t, it has to be that he was chosen by i, *i.e.* $d_{ijt} = 1$.

Table 1.15: Amount paid back: estimation results

	P1	P2	P3	P4
$S \underset{t-1}{\dashrightarrow} R$	24.21	13.3	12.04	17.71
	(5.88)***	(4.8)***	(4.78)**	(4.5)***
$S \underset{t-1}{\dashleftarrow} R$	1.69	3.33	2.89	-1.52
	(5.99)	(4.65)	(4.62)	(4.64)
$S \underset{t}{\dashleftarrow} R$	32.68	26.25	25.67	19.45
	(5.16)***	(4.56)***	(4.54)***	(4.6)***
t_{jit}	.	1.83	.14	.
		(.09)***	(.28)	
t_{jit}^2	.	.	.01	.
			(.002)***	
period 3	-3.95	-7.24	-6.14	-14.38
	(4.34)	(3.77)*	(3.77)	(9.06)
period 4	-10.14	-12.57	-12.16	-3.95
	(4.93)**	(3.84)***	(3.86)***	(7.98)
period 5	-28.42	-29.73	-29.08	-1.17
	(6.3)***	(5.6)***	(5.63)***	(7.93)
period 6	-74.83	-65	-65.22	-4.96
	(7.82)***	(7.06)***	(6.99)***	(9.63)
$t_{ji}\cdot$ period 2	.	.	.	2.11
				(.11)***
$t_{ji}\cdot$ period 3	.	.	.	2.18
				(.12)***
$t_{ji}\cdot$ period 4	.	.	.	2.01
				(.11)***
$t_{ji}\cdot$ period 5	.	.	.	1.77
				(.13)***
$t_{ji}\cdot$ period 6	.	.	.	1.49
				(.17)***
Const.	35.02	-109.44	-64.35	-55.84
	(68.58)	(57.85)*	(58.4)	(57.81)
Obs.	2540	2540	2540	2540
R^2	.22	.42	.43	.42

Note: Variables refer to periods as indicated by the subscripts. S is sender and R is the receiver, t_{jit} is the transfer R received in t. Control variables are: gender, age, siblings, and nationality of sender and receiver, dummies for session and treatments.

Table 1.15 presents results for four specifications. The labels S for sender and R for receiver remain unchanged, even though it is now the receiver to take action. Note that by default, the sender must have chosen the receiver in period t, otherwise the receiver does not make a move. Thus, either S and R formed a pair in t (if $(S \leftarrow\!-\!- R)_t$ is one), or S chose R but not vice versa.

Specification P1 forms the starting point for the analysis. We find that payback is on average about 24 units higher when receiving a transfer from the same subject for the second time. This is evidence supporting hypothesis RIF 3 and is evidence against hypothesis DLT 3. We also find that a subject pays back about 33 units more if he also sent a transfer to this subject in the same round. This effect is not directly associated to one of our hypotheses. Notice that sending a transfer to this subject in the previous round has no significant effect (see coefficient of $S \underset{t-1}{\leftarrow\!-\!-} R$).

In specification P2 of Table 1.15 one can find support for hypothesis RCP 5. The amount received increases the amount paid back. In particular, the coefficient of the transfer received is 1.8. Together with the insignificant constant this shows that on average for each token sent, which is multiplied by three, subjects pay back 1.8, or, equivalently 60 percent of the amount received. The functional relation between the transfer received and the amount paid back is not linear, as can be seen from the positive and significant coefficient of the variable t_{ji}^2 in specification P3. However, the curvature does not have a substantial impact in the range 20 to 100 where 90 percent of the transfers can be found, and a linear relation can be seen as a good approximation.

One might be pressed to interpret the decreasing period dummies in P1 to P3 as evidence for hypothesis RTN 2 that payback is lower in the last periods. However, when we add interaction terms between period number and transfer received (see P4) we find that the period dummies become insignificant. Instead, now we find a significant decrease in the ratio returned from 70 percent in period 4 to 50 percent in period 6. While transfers are lower in final periods (as shown in previous section) less can only be paid back. Nevertheless, the decline in payback can be explained only by a declining return ratio. In other words, the rate of the decline of payback is stronger than that of the decline in transfers.

Table 1.16 summarizes the main findings of this section.

Table 1.16: Amount paid back: summary of findings

	...*the amount paid back is higher*	*evidence*
RCP 5	if the transfer received was high	yes
RIF 3	if received for consecutive periods	yes
	...*the amount paid back is lower*	
DLT 3	if received for consecutive periods	no
RTN 2	in the last period	no
additional findings		
pay back is higher if transfer sent to same player		
the return ratio is lower in last periods		

1.7 Final Remark

This paper analyzes the determinants of trust and trustworthiness in an experiment where trust can emerge as the result of repeated interaction between individuals. We add an element of choice to the setting of a repeated trust game, in that players have the opportunity to choose among four players. For each opponent, players see information such as age, nationality and gender. The influence of four different behavioural and learning theories is looked at: directional learning, reinforcement learning, reciprocity and rationality. The econometric analysis goes along the three stages of the game: choice, transfer and return, controlling for confounding factors. It sheds light on the behavioural motives behind each decision. The low degree of formalization and a certain degree of observational equivalence makes a clear discrimination between the competing approaches impossible. While it is not possible to attribute the entire playing behaviour to a single type, at each decision several motives seem to influence the decisions taken, some being of higher explanatory power than others. It was shown that a mixture of several motives is at play at each stage of the game. In the same way as rationality does not offer a satisfying explanation for the behaviour of the players, none of the alternative motives such as reinforcement or reciprocity is able to capture all facets of the observed behaviour.

CHAPTER 2

Choosing Who to Trust

2.1 Introduction

Most decisions in daily life involve a discrete choice. Consumers decide to buy a second-hand car from one particular dealer and not from another, holiday makers choose one hotel among many at their destination, *etc.* Many of these decisions involve a considerable degree of uncertainty regarding the characteristics of the choices and consequences of decisions. Because markets are imperfect and it is impossible to incorporate all contingencies into contracts, people often have to rely on their believes and attitudes when engaging into economic interactions. For example, at a certain point a consumer will just have to *trust* the word of the car dealer that the car he is about to buy is working properly. While economic theory is mostly about the interaction of anonymous agents, in reality people pay attention to identifying characteristics. Trust, broadly defined as the the conscious engagement in actions that increase one's own vulnerability (Kollock, 1994), is likely to depend, among others, on cultural background. Differences in the way daily life is organized in societies are likely to induce differences in the level of trust people place in strangers.[1] Varying attitudes towards trust are also likely to be the result of gender differences. The local authorities of Mexico City, for example, have introduced all female teams of traffic police in an attempt to reduce the incidence of bribes, because they trust women to be less corrupt than men (Buchan et al., 2004).

1 For evidence on this, see e.g. Croson and Buchan (1999).

As the list of examples above suggests, many real life situations often display a degree of complexity that is difficult to capture in a simple and tractable model. But to understand human behaviour it is important to identify the reasons why individuals decide the way they do; why the consumer trusts one car dealer more than any other. What adds to the difficulty is that observable data, even if collected at the individual level, very rarely includes information about all options faced by the individual. But if something is to be said about why an individual prefers one choice over another, it is essential to know the characteristics of all choices.

The experiment analyzed in this part of the thesis offers the possibility to shed some light on the determinants of trust in a discrete choice setting. Individuals were matched in groups of five players to play the repeated trust game. Every period each player could choose one among the four other players in his group to play the game. Each player had information about nationality, gender, age and the number of siblings of all other players in his or her group. This trust game was repeated for six periods before groups were reshuffled.[2] The experimental setup allows to track the presence of trust and trustworthiness along three dimensions. The first and probably most important dimension is choice. Because the division of the surplus in the trust game depends on the willingness of the receiver to reward the sender, the sender should think carefully which player to trust. Having made that decision the next dimension of trust is how much the sender wants to transfer to the receiver. Lastly, the amount that the receiver returns to the sender is a measure of his trustworthiness. At this stage the initial believes of the sender about the trustworthiness of the receiver are either confirmed or not.

The information collected in this setup is exceptionally rich because in addition to the individual's characteristics, the characteristics of the entire choice set are observable to the econometrician. Hence this is one of the rare cases in which it is possible to apply the choice framework developed by McFadden (1973), more specifically, the conditional logit model to determine what leads subjects to trust one player but not another.

This chapter analyzes which of the observable characteristics of the players influenced their choice of a player, how these determinants evolved in the course of the game and if any significant deviation from randomness is payoff relevant. The analysis shows that male players have a strong preference for

2 For details about the experimental setup the reader is referred to the description of the experiment in Section 1.3 on page 12 and to the Appendix A.1. Players could decide if they wanted to make a positive transfer to any of their fellow players or they could decide not to make a transfer at all. In the subsequent analysis only cases in which players actually decides to play will be considered. The analysis here only includes the first four treatments where individuals had free choice.

transferring points to female players, while female players do not discriminate in such a way. Interestingly, this behaviour is not payoff relevant, in the sense that women do not reciprocate transfers more than men by making higher returns. Along the lines of nationality, this chapter discovers a significant preference of northern players to play with players from the same region, discriminating against southern players. The evolution of this discrimination and its payoff relevance is subject to an in-depth analysis in Chapter 3.

2.2 An Application of the Conditional Logit Model

2.2.1 Specification

The multinomial choice model is best motivated by the random utility model. Call U_{ij} the utility of player i if he decided to choose alternative j and define

$$d_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses } j, \\ 0 & \text{otherwise.} \end{cases}$$

In the experiment, each player has four mutually exclusive and exhaustive choices, $j = \{1, 2, 3, 4\}$. The random utility model relates the characteristics of individual i and of the choice j to the utility level U_{ij}. Hence:

$$U_{ij} = \alpha W_i + \gamma Z_j + \epsilon_{ij}, \tag{2.1}$$

where W_i contains i-specific information such as nationality and gender of player i and Z_j contains information that varies across choices. It includes all information that player i has about his fellow players j (gender, nationality, age, siblings). The errors ϵ_{ij} are assumed to be independently distributed both across i and j.

In general, in the conditional choice model, it is not possible to identify coefficients of variables that do not vary within groups.[3] This does not mean, however, that individual effects are not accounted for: i–specific characteristics just do not affect the *relative* probabilities of choice. Effects of attributes of a player i such as gender and nationality can be identified if they are interacted with choice varying characteristics. By doing so, and given some variation across choices, it is possible to identify the effect of i-specific characteristics relative to the characteristics of j. To this end, define the following interacted variables:

[3] In this case a group refers to the one out of four players choice faced by each individual in each period of each treatment.

mm_{ij} takes value one if both i and j are male and zero otherwise,
ff_{ij} takes value one if both i and j are female and zero otherwise,
NN_{ij} takes value one if both i and j are from North and zero otherwise,
SS_{ij} takes value one if both i and j are from South and zero otherwise.

Note that these four variables are a complete set of interactions of gender and nationality.[4] For example, if i is male, mm_{ij} may take values 0 or 1, but ff_{ij} is obviously zero. In addition, define the following two variables of j-specific characteristics:

age_j the age of player j,
sib_j the number of siblings of player j.

Define the vector $X_{ij} = (mm_{ij}, ff_{ij}, NN_{ij}, SS_{ij}, age_j, sib_j)$ and a corresponding parameter vector β, then the random utility model becomes:

$$U_{ij} = \beta X_{ij} + \epsilon_{ij}. \tag{2.2}$$

Player i chooses player j if this yields highest utility. Hence,

$$\mathrm{P}(d_{ij} = 1) = \mathrm{P}(U_{ij} > U_{ik})\ \forall\, k \neq j.$$

Assume further that the errors are distributed independently identically across i and j and with $F(\epsilon_{ij}) = \exp(-e^{-\epsilon_{ij}})$, so that the model takes the form of the conditional logit model (Maddala, 1993, pp. 60–61). Then the probability of a choice j is

$$\mathrm{P}(d_{ij} = 1) = \frac{e^{\beta \mathbf{X}_{ij}}}{\sum_{j=1}^{4} e^{\beta \mathbf{X}_{ij}}}. \tag{2.3}$$

which is the standard McFadden (1973) conditional logit specification.[5]

The log-likelihood function $\log L(\beta)$ is the sum over all such probabilities for all $J = 4$ choices, for all individuals i, for all periods p and treatments t:

$$\log L(\boldsymbol{\beta}) = \sum_{T=1}^{T}\sum_{p=1}^{P}\sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij}^{pt} \log \mathrm{P}(d_{ij}^{pt} = 1),$$

4 For the analysis of nationality the participants were grouped into two geographical regions North and South, see also the discussion in Chapter 3. Table 3.1 on page 51 lists the countries, their average latitude and number of participants in the experiment.

5 Notice that throughout the experiment subjects were constrained in the number of alternative choices, and the choice set was complete and mutually exclusive by design. Hence, the underlying assumption that derives from the particular error structure of the conditional logit model and which poses a problem for many of its applications, the assumed independency of irrelevant alternatives (IIA), is fulfilled by construction.

where $T = 4$, $P = 6$, $N = 110$, which yields a total of 2640 units facing a four choices decision problem. In 166 cases players decided not to play the game so that 2474 units remain, or, equivalently, 9896 observations.[6]

The goodness of fit of the model can be assessed by the share of correct predictions made by the model. This is the percentage of correct choices predicted by the model disregarding the contribution of the error component ϵ_{ij}. The predicted choice is the choice that was attributed the highest probability. More specifically, consider the choice in a specific treatment t and a specific period p. Define:

$$\hat{j}_i = \max_j \hat{P}_{ij} \text{ , where } \hat{P}_{ij} = \frac{e^{\hat{\beta}\mathbf{X}_{ij}}}{\sum_{j=1}^{J} e^{\hat{\beta}\mathbf{X}_{ij}}}.$$

Denote the actual choice of individual i as

$$j_i^* = \max_j d_{ij}.$$

The fit for a particular period/treatment combination pt is measured by:

$$F^{pt} = \frac{1}{N}\sum_{i}^{N} \mathbf{1}\{j_i^* = \hat{j}_i\},$$

where $\mathbf{1}\{...\}$ is an indicator function that takes value one if the expression inside the parentheses is true and zero otherwise.

2.2.2 Estimation Results

Table 2.1 reports estimation results for equation (2.3) for three sub-samples of the data. The first specification C1 uses all observations for the first four treatments, the second column (specification C2) uses only the data from the first period of the first treatment, while the last specification C3 is restricted to the first period of all four treatments.

The results show a clear discrimination taking place along the lines of nationality. Over all periods and treatments (specification C1), northern players have a strong preference for other northern players. On average, the odds of a northern player choosing a northern player instead of a southern player are 1.38. Southerners also have a preference for northern players, although the effects are not significant. Still, the odds of a southerner choosing another

6 Contrary to the analysis in Chapter 1 the analysis does not include the last free choice treatment, where subjects just had completed the predetermined treatment of forced interactions.

player from South are just 0.91. In the first period of the first treatment (specification C2), where much less observations are available, the odds for nationality go into the same direction, but they are not statistically significant. If the sample is restricted to the first periods of all treatments (specification C3) the preference of North for North is reconfirmed.

But nationality is not the only dimension along which discrimination takes place. The effect of gender is also clearly visible. Male players have a strong preference to contact female players throughout all specifications. This effect is particular strong in the first period of the first treatment (C2). In this case, the odds of a male player choosing another male player are just 0.53, compared to 0.76 overall. The effect is borderline significant even tough this specification includes very few observations.

Other demographic variables also influence choice. Overall, the effect of age is positive and significant. The odds of a player being chosen increases with his age, but the magnitude is very small. The significance is not confirmed in the sub-samples of the data, where the odds of choosing a player are unaffected by his age. Interestingly, there is a consistent pattern when it comes to the influence of siblings on choice. Compared to the excluded case that the player chosen is a single child, the odds of a player being chosen increase with the number of siblings that he or she has and peak at three siblings.[7]

The share of correct predictions in both specifications is well above the random prediction of 0.25. In particular in the specification of the initial choice (first period, first treatment), where no playing experience influences the players' choice, 43 percent can be predicted by this simple model.

Notice that the estimates presented in specification C1 in Table 2.1 include estimates of all periods. However, in the second period of the game players have already made experiences with their choice and might revise their decision. It is therefore interesting to see how the estimates evolve in the course of the game. Figure 2.1 plots the first five coefficients from equation (2.2) estimated for each period p, together with the corresponding standard error.

The evolutions of the coefficients of the gender variables confirms the general picture that emerged from the table above. There is a strong preference of males to play with females at early stages in the game (see the first panel of Figure 2.1). Interestingly, in the second period there is a pronounced revision of this preference when males are indifferent in their choice, before the preference starts building up again in the course of the game.

7 The cases of $sib_j = 3$ and $sib_j = 4+$ include 5 and 7 subjects only, so the peak at $sib_j = 3$ should not be over-interpreted. In results not reported here, interactions of the number siblings of i and the number of siblings of j were also included in the estimations. While this increased slightly the predictive power of the models, no consistent pattern emerged. The same is true for age.

Table 2.1: Conditional Logit estimation results

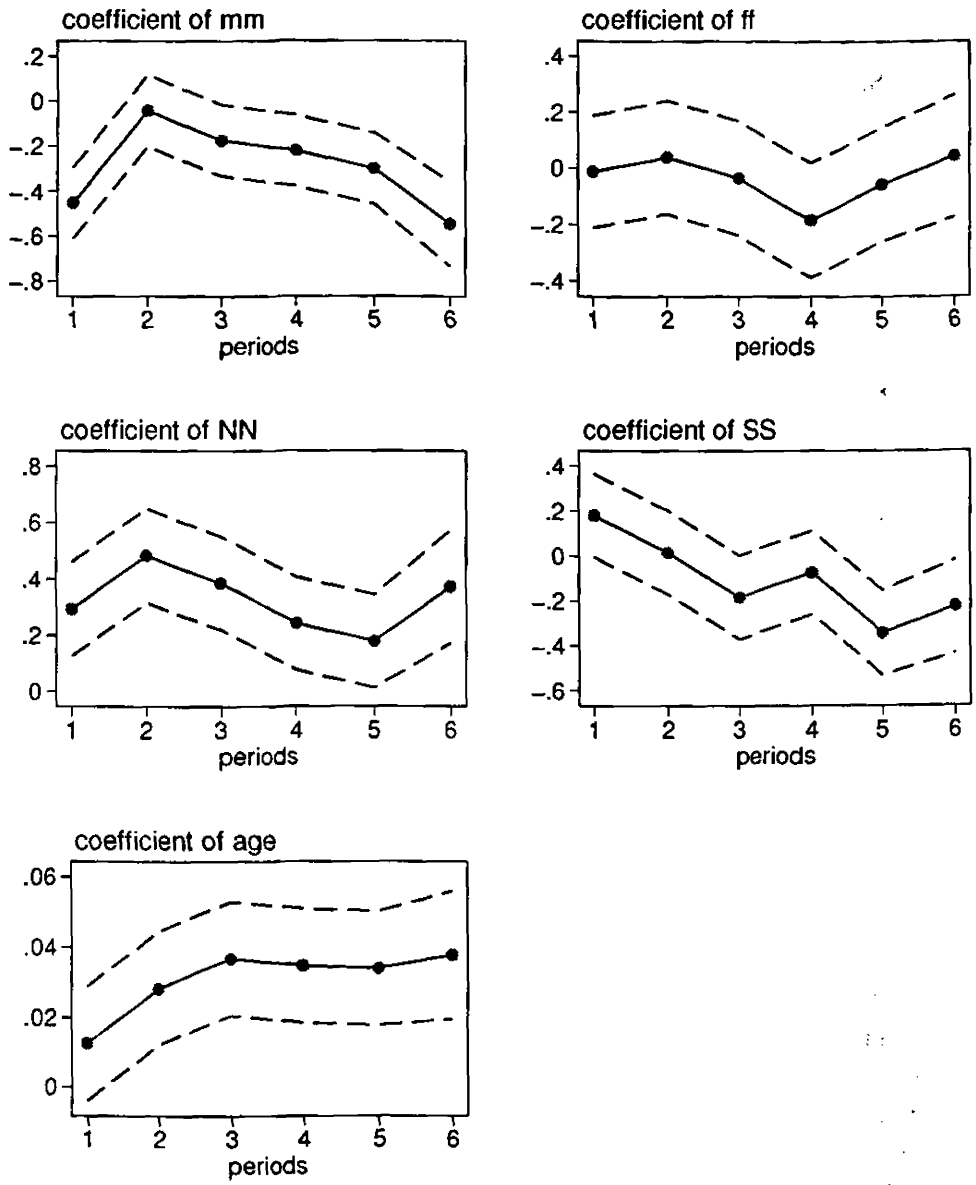
	C1	C2	C3
NN_{ij}	1.38 (.09)***	1.40 (.54)	1.34 (.22)*
SS_{ij}	0.91 (.07)	0.67 (.28)	1.20 (.21)
mm_{ij}	0.76 (.05)***	0.53 (.18)*	0.64 (.10)***
ff_{ij}	0.97 (.08)	1.36 (.52)	0.99 (.19)
age_j	1.03 (.01)***	0.95 (.03)	1.01 (.02)
$sib_j = 1$	1.09 (.08)	1.59 (.67)	1.11 (.21)
$sib_j = 2$	1.34 (.12)***	2.34 (1.07)*	1.17 (.24)
$sib_j = 3$	1.32 (.17)**	4.41 (2.91)**	2.29 (.68)***
$sib_j = 4+$	1.60 (.19)***	1.18 (.86)	1.39 (.40)
Obs.	9896	420	1716
Correct Predictions	0.31	0.43	0.35
treatments T included	1,2,3,4	1	1,2,3,4
periods p included	all	1	1

Note: The table reports odds ratios, standard errors in parentheses. $*$, $**$ and $***$ is the significance of a a test at the 10, 5 and 1 percent level respectively that the odds ratios are different from one. Correct predictions indicates share of cases in which the maximal predicted probability coincides with the actual choice of i.

There is no such effect for female players, none of the estimates is significantly different from zero (see second panel of Figure 2.1).

The preference of northern players to choose northern counterparts is positive and significant throughout the game. For southerners a change in preference can be observed. As the fourth panel in Figure 2.1 shows, they start off with a slight preference for South, but end up choosing more players from North as the game evolves. As will become clear in the Chapter 3, this is directly related to differences in playing behaviour between South and North.

Figure 2.1: Coefficients over periods



Note: Scaling differs from graph to graph. Dashed lines are 90% confidence intervals

The results form this section can be summarized as follows.

- Men enter the game with a strong preference to contact women. While this preference weakens considerably in the following period, it still turns out to be significant in subsequent periods.

- Women instead do not discriminate based on the gender of their counterparts.
- Northern players discriminate against southern players in that they are more likely to choose someone from North throughout the game.
- While southern players have an overall tendency to prefer northern players, this is only at borderline significance and it builds up in the course of the game.

2.3 Does it Pay Off to Discriminate?

The question that emerges is if the observed discrimination pays off, *i.e.* if it is rational for men to contact women instead of men.[8] If women are more trustworthy than men, it is rational for men not only to contact them more often but also to make higher transfers, because they can expect to get a high return. In order to verify this, the gender discrimination will be analyzed in terms of the transfers sent to other players and how much is returned.

Indeed, male players make higher transfers, but they do so to both men and women. The left hand side of Table 2.2 reports the average transfers made by male and female senders to their male and female counterparts. The first observation is that, both, over all treatments but also in the first period of the first treatment, men transfer more points than women. Second, over all four treatments, the raw data also suggests that men receive more than women. Notice however, that in the first period of the first treatment, females send substantially less to men than to women (33 *vs.* 43).

The relevant question in this context is not if men *in general* trust more than women, but if men trust women more than men.[9] To test the significance of differences in transfers sent between genders and control for confounding factors, the following framework is used. Denote by t_i the transfer sent by player i to player j and define an indicator variable fem_j that takes value one if the receiver j is female. The following regression is estimated:

$$t_i = a + \alpha fem_j + \delta X + u_i. \tag{2.4}$$

The matrix X contains the control variables age, nationality and siblings of the sender and the receiver. If a sender sends more to a female receiver, the

8 The discrimination based on nationality is subject to the analysis in Chapter 3 and will not be discussed here.

9 Men, in general, make higher transfers than women, even controlling for confounding factors (results not reported here).

estimated coefficient $\hat{\alpha}$ should be different from zero. The right hand side of Table 2.2 reports results for the regression in equation (2.4) by sender groups.[10] Controlling for observational characteristics, none of the above differences is significant. A first result of this section is that, judged on transfers received, male and female players do not differ significantly. If anything, female players make low transfers to male players at early stages in the game. Interestingly, while male players strongly prefer female players as partners, they do not trust them more in the sense that they make higher transfers to them.

Table 2.2: Gender effects on transfers received

		male receiver	female receiver	α	s.e	t-val.
$T = 1, 2, 3, 4$	male sender	83	80	-1.49	2.54	-0.59
	(obs.)	(927)	(628)			
	female sender	77	73	-2.57	3.10	-0.83
	(obs.)	(601)	(318)			
$p = 1, T = 1$	male sender	56	54	-5.46	10.12	-0.54
	(obs.)	(36)	(30)			
	female sender	33	43	7.93	9.53	0.83
	(obs.)	(21)	(18)			

Note: The left hand side of the table reports the average transfer to male and female receivers by sender groups, for the first four treatments and for the first period, first treatment separately. The right hand side of the table reports, by sender groups, results of the regression $t_i = a + \alpha fem_j + \delta X + u_i$, where fem_j takes value 1 if the receiver j is female. X is a matrix of control variables including age, nationality and siblings of sender and receiver. u_i is an error component. Standard errors in the first row corrected for multiple observations per individual. t-value reported for the test $\alpha = 0$.

But does the return behaviour give a reason for males to discriminate in favour of females? To investigate this question, a similar analysis is carried out for the amount returned. Define the ratio returned to a sender i as $r_i = G_i/(3 \cdot t_i)$ where t_i is the original transfer and G_i is what receiver j returned to player i. Table 2.3 reports on the left hand side, by gender of sender and receiver, the ratios returned. In the table, rows are senders, columns are receivers. It appears as if, over all treatments, both sexes tend to return more to the same sex than to the opposite sex. To see if these differences are significant, the following regression (analogous to equation (2.4)) is estimated.

10 Standard errors are corrected for the fact that in the upper row of the table repeated observations for players were used, see Moulton (1986).

In the regression:

$$r_i = b + \beta fem_j + \delta X + e_i \tag{2.5}$$

one can test the significance of gender effects of the return ratio by testing if the coefficient $\beta = 0$. The right hand side of Table 2.3 reports this by sender groups. Over all treatments the differences are not significant. However, in the first period of the first treatment, despite the fact that there are only so few observations, it can be seen that males get back significantly less from females than from males (0.47 vs. 0.55). Notice that the average return ratio from males to females instead is 0.63, much higher than any other value observed. Clearly in their first contact, the two sexes treat each other very unequally.

Table 2.3: Gender effects on the ratio returned

		male receiver	female receiver	β	s.e	t-val.
$T = 1,2,3,4$	male sender	0.56	0.53	-0.02	0.03	-0.64
	(obs.)	(927)	(628)			
	female sender	0.53	0.58	0.01	0.03	0.54
	(obs.)	(601)	(318)			
$p = 1, T = 1$	male sender	0.55	0.47	-0.17	0.06	-2.59
	(obs.)	(36)	(30)			
	female sender	0.63	0.48	-0.15	0.09	-1.77
	(obs.)	(21)	(18)			

Note: The left hand side of the table reports the average ratio returned by male and female receivers to sender groups, for the first four treatments and for the first period, first treatment separately. The right hand side of the table reports, by sender groups, results of the regression $r_i = b + \beta fem_j + \delta X + e_i$, where fem_j takes value 1 if the receiver j is female. X is a matrix of control variables including age, nationality and siblings of sender and receiver. e_i is an error component. Standard errors in the first row corrected for multiple observations per individual. t-value reported for the test $\beta = 0$.

2.4 Discussion

From the results above one can conclude that the initial trust put in women by men in contacting them more often is not followed by higher transfers, and, more importantly, is not reciprocated by higher returns.

One possible interpretation is that men initially believed that women are more trustworthy and hence expected them to return more. But after realizing that they receive less from women in general, and that women do not meet their own standards in the ratio returned (in particular they return far less than men do to the opposite sex) they revise their beliefs and switch to other players. This explanation is supported by the increased probability of males in choosing male partners in the second period, as seen in the first panel of Figure 2.1.

Had subjects read the literature on gender and trust and behaved accordingly, the picture would have been less clear. The literature does not offer a conclusive answer to which gender is more trusting and which is more trustworthy. In an international study of the trust game, Croson and Buchan (1999) find that women do not give significantly more than men but that they are significantly more reciprocal in their return behaviour. Here it would have been rational to transfer more to women. However, in their study, the gender of the players is not made public among participants. Evidence for significant gender differences when the sex of the opponent is known is provided by Eckel and Wilson (2003). But in their simplified version of the trust game, they find that women are more likely to trust than men, but do not observe any differences in payback behaviour.[11] More related to the present study is Buchan et al. (2004), who carefully investigate how information about the opponent's sex affects trust and trustworthiness. In a standard one shot trust game, they vary the information that sender and receiver have about their (matched) opponent. While they find that men send significantly more than women throughout their treatments (as in this study), in their treatment where the sex of both sender and receiver is common knowledge they do not find any additional effect, neither in transfers sent nor in the amount returned.

These findings are consistent with the result of study. The common belief that women are more trustworthy than men is not confirmed in a setting where both partners have information about their opponents' sex.

2.5 Concluding Remark

In a setting of the repeated trust game with an element of choice, where players had information about their opponent's gender, age, and nationality, the choice of a player was found to be influenced by these characteristics. Using the conditional choice model of McFadden (1973) the analysis reveals that, especially upon first contact, men are more likely to contact women than other men. The second important finding is that subjects from northern

11 In their study the first mover could not decide *how much* to transfer to the receiver, but only if the total amount should be transferred or not.

European countries are more likely to choose players from the same region. However, the discrimination that takes place based on the gender information is not payoff relevant: women do not reciprocate transfers with higher returns.

The effect of gender on trust and trustworthiness as investigated by the literature does not yet yield a conclusive picture. While there is some evidence that women are trusted more, this expectation cannot always be confirmed. Outside the experimental laboratory, these differences are likely to be more complex, and perhaps largely contextual. Whether female car dealers are more successful than their male counterparts remains an open question.

CHAPTER 3

Trust and Trustworthiness Among Europeans: South – North Comparison[1]

3.1 Introduction

Most economic interactions are preceded by a stage in which agents select partners. Entrepreneurs select their counterparts for a partnership, firms select suppliers, consumers select retailers and employers choose workers from pools of applicants. The initial choice of a partner as well as the decision about the volume of activity to a large extent depends on the agent's beliefs about the prospects of building trust and reciprocity with potential partners. If the interaction takes place repeatedly, experience will play a role as well. Selectors are expected to return to those partners who proved to be trustworthy, and avoid those who failed to reciprocate. In a global environment where economic interactions go across countries and cultures, national diversity may have a substantial impact on agents' initial beliefs regarding partners as well as on the evolution of their interaction over multiple transactions.

In this paper we report on experimental results that describe the impact of cultural diversity on agents' choices of partners as well as on the outcomes of economic interactions. Our subject pool involves participants from different European nationalities. Dividing the continent into two regions our objective is to compare subjects' perceptions about trust and reciprocity between northern and southern Europe by studying subjects' choices of partners and the volume of economic activity.

1 This chapter is joint work with Andrea Ichino, Karl Schlag and Eyal Winter.

The issues of trust and reciprocity in economic interactions have been given a considerable attention by the recent literature in experimental economics. Berg et al. (1995) addressed these issues by designing a trust game experiment. The standard format of a trust game involves two players. The "sender" who is assigned an amount of money x by the experimenter decides on a transfer $0 \leq t \leq x$ to be made to the "receiver", who will receive three times the amount of this transfer, i.e., if the sender concedes the amount t to the receiver, then the latter receives $3t$ (while the sender loses just t). Following the transfer made by the sender, the receiver has to decide how much she wants to return. The amount that the receiver decides to return is denoted by $g \in [0, 3t]$ and is equal to what the sender gets back. While the unique Nash equilibrium prediction of the game is for the receiver to make zero payback and therefore for the sender to make no transfer at all, Berg et al. (1995) found that senders did make considerable transfers, which are backed by substantial paybacks. Among other papers that study subjects' behavior in this trust game is Buchan et al. (2000), which involves a comparison across different countries including the US, China, Japan and Korea focusing on the effect of preliminary discussions within groups on behavior in the trust game.

Our framework differs from this strand of literature in three major aspects. First, we are not interested in differences across countries when subjects interact with partners of the same nationality. We are instead interested in differences across countries when subjects from different nationalities jointly play together our version of the trust game. Second, to highlight the role of the choice of partner in real settings we have allowed participants to choose the partner to whom they make a transfer. Finally, we have designed a dynamic version of the trust game to allow trust and reciprocity to be built up and to enable us to study the evolution of trust in our multi-cultural framework. Our version of the trust game will be described in greater detail in Section 3.2.

Somewhat more related to our framework is Fershtman and Gneezy (2001) which reports results on a one shot trust game played between Ashkenazi (Jews of European descent) and Sephardi (Jews of Middle Eastern origin) Israelis. They found that Sephardi subjects were discriminated against in the amount of transfers they received although their payback behavior wasn't different from that of their Ashkenazi counterparts. In contrast to their framework in which matching was fixed and the interaction involved a one shot game, in our framework each subject can act both as a sender and a receiver; subjects choose their partner and interact repeatedly within the same group. These features will allow us not only to detect discrimination but also to go more deeply into its roots by analyzing the way it evolves over different periods.

We have conducted our experiment in an environment where a major role of nationality is least expected. Our subject pool involves Ph.D. students at the European University Institute (EUI) in Florence. The EUI whose main objective is to provide advanced academic training to Ph.D. students in a European perspective, attracts young intellectuals from EU member countries with substantial international exposure and with a typical fluency in at least three European languages. If the role of nationality within this group is strong, we would expect it to be even stronger among the general population of Europe.

Our results, presented in Sections 3.3, 3.4 and 3.5, indicate discrimination against South in terms of number of contacts, carried out mainly by northern subjects. However, the most interesting finding is the fact that this discrimination builds up rather than dying out with experience. More than for not being trustworthy (i.e. having a low propensity to reciprocate by making a generous payback for a transfer received), Southern Europeans are being punished for their own low level of trust (i.e. having a low propensity to contact another player with a generous transfer), and for this reason ends up leaving the game with lower payoffs.

As discussed in the concluding Section 3.6, we find these results particularly striking because of the international exposure of our group of subjects. We interpret these results as an indication that cultural differences in standards regarding trust and reciprocity, possibly related in our case to the stage of development or to the role of the family in the two regions, may be sufficiently robust to persist even when individuals change their original habitat.

3.2 The Design

The design of our experiment is described extensively in Appendix A.1. Here we limit ourselves to a summary of its most important features. We conducted three sessions with a total of 110 participants hired among EUI Ph.D. students from different European countries. Upon entry, subjects were asked to fill in a form in which they had to specify their nationality in addition to other bits of personal information (gender, age and number of siblings) that still allowed their identity to be kept anonymous. This was mainly done in order to blur the fact that our interest lies with the issue of nationality. In each session subjects played six treatments in each of which they were assigned randomly to a group of five players. At the beginning of each treatment the personal information about the other players was made public within the group.

In the first four treatments subjects were allowed to choose without restrictions the partner with whom they wanted to interact among the four subjects in their group. Every one of these treatments involved 6 periods with the following structure: At the beginning of the period, each player received an endowment of 100 points, equivalent to 0.35 Euros. Subjects were then given the opportunity to transfer any part of the initial endowment to a single player of their choice within the group.[2] If a sender made a transfer of t to a receiver she received $3t$. Then each subject who received a transfer had an option to return back any part of it to the person who made him the transfer and the period ended. All decisions were made via computer terminals. At the end of each period a subject saw on the screen only the actions and payoffs of the interactions in which she had been involved (i.e. the one in which she had been a sender if she had transferred a positive amount to some player, and the ones in which she had been a receiver if she had received a transfer from one or more other players). Thus, subjects did not know at the end of a period what had happened between other pairs of players.

The fifth treatment differed from the previous four because subjects were randomly matched to another player at the beginning of each period. Thus, they did not have the possibility to choose a partner and could only send to the player randomly assigned to them. The sixth and final treatment was instead identical to the first four.

Most of our analysis will involve the first four treatments, since the last two treatments are distorted by the absence of free choice in the fifth treatment. However, we will also look at how imposing a partner in the fifth treatment changes the behavior of subjects in the last treatment in which the choice of partner is again free.

In reporting the results we will refer to two characteristics of players: "trust" and "trustworthiness". Trust concerns sending behavior. It refers to the propensity of a player to contact another player and to make high transfers, which we interpret as a propensity to trust the receiver to reciprocate.[3] By looking at the aggregate data within each region we will provide analysis regarding the extent to which region H trusts region K, where $H, K \in$ {North, South}. This will be done by looking at the propensity by which players from region H choose to make a transfer to players of region K as well as the amount of transfers they make. Trustworthiness stands for the tendency of a player to reciprocate by making a generous payback for a transfer he/she received. At the regional level it will be measured by the

2 Note that choosing no partner was possible, in which case the transfer was equal to zero.

3 Note that, in our framework, lack of trust can emerge either because senders assign a small probability to the event that their partner will reciprocate or because senders are risk averse. The distinction between these two possible reasons for not trusting others is outside the scope of this paper.

average return ratio, i.e. what receivers return to senders as a fraction of what they have received. The precise statistics will be explained later. We provide the analysis at the regional level and not at the country level as we fail to have sufficiently many observations for each country pair separately.

3.3 Results

The evidence provided in this paper is based on the aggregation of countries in two regions (South, North) according to their average geographical latitude. Table 3.1 lists the countries represented in each region, the average latitude (in degrees) of each country and the number of subjects per country.

Table 3.1: Nationalities: frequencies and average latitude

country	av. latitude	participants
Southern countries		
Greece	39	9
Portugal	39.3	1
Spain	40	11
Italy	42.5	17
France	46	12
Northern countries		
Austria	47.2	6
Belgium	50.5	5
Germany	51	16
Poland	52	3
Netherlands	52.3	8
Ireland	53	5
United Kingdom	54	8
Denmark	56	3
Sweden	62	4
Finland	64	2

Note: For average latitude see CIA (2003).

Table 3.2 provides some general descriptive statistics based on the following notation. The variable f_i^N is the frequency of northern players seen by sender i in her group. With five randomly selected players in each group a sender sees four players and thus the variable takes the following values: $f_i^N \in \{0, 0.25, 0.5, 0.75, 1\}$. For each of these values we have a column in

Table 3.2. Notice that whenever this variable takes value 0 the sender faces only players from South and hence cannot choose a player from North. The converse is true if $f_i^N = 1$.

We denote the average frequency of zero transfers (i.e. no choice of partner) for senders of region $H \in \{N, S\}$ by $Z_H(t = 0)$. These frequencies are reported in the first row of Table 3.2 for each value of f_i^N. Interestingly, with the exception of the shift from $f_i^N = 0$ to $f_i^N = 0.25$, $Z_N(t = 0)$ decreases with the fraction of northern players seen by the sender, indicating that on average North are[4] more willing to trust when a larger number of interactions with North is possible. The opposite pattern prevails instead for South since $Z_S(t = 0)$ increases with f_i^N.

Table 3.2: Descriptive statistics

	f_i^N					
	0	0.25	0.5	0.75	1	average
$Z_N(t=0)$	0.07	0.08	0.07	0.04	0.04	0.06
$Z_S(t=0)$	0.03	0.04	0.05	0.10	0.11	0.07
t_N	75	78	80	80	77	79
t_S	60	77	66	67	71	69
r_N	n.a.	0.52	0.54	0.57	0.60	0.56
r_S	0.56	0.56	0.50	0.55	n.a.	0.54
$\pi_{N,send}$	177	147	154	163	173	158
$\pi_{S,send}$	132	163	136	147	149	146
$\pi_{N,rec}$	137	116	106	96	82	105
$\pi_{S,rec}$	88	74	92	86	98	87
$\pi_{N,total}$	314	264	260	259	255	262
$\pi_{S,total}$	220	237	227	233	247	233

Note: f_i^N is the frequency of northern players seen by sender i in her group. $Z_K(t = 0)$ is the fraction of zero transfers for senders of region K. t_K is the average transfers sent by region K. r_K is the average return ratio chosen by receivers of region K. This figure is not available for northerners (southerners) in the cases in which only southerners (northerners) are seen by senders. The payoff for a sender of region K from making a transfer is defined as $\pi_{K,send} = 100 - t + g$ where g is what the sender gets back. The payoff for a receiver of region K from receiving transfers is defined as $\pi_{K,rec}$= (sum of total amounts received by other players − sum total amounts returned to other players). The total payoff is the sum $\pi_{K,total} = \pi_{K,send} + \pi_{K,rec}$. K is equal to N or S denoting North and South respectively.

4 In the sequel we will use the terms "South" and "North" to refer to the plurality of subjects from the two regions. Thus, "South" and "North" will be short for "Southerners" and "Northerners".

The average transfers by senders of the two regions are denoted by t_N and t_S and are displayed in the second row of the table. For all values of f_i^N northern senders transfer more tokens than southern senders, which indicates that North have a larger propensity to trust. In the first column of the table it is also worth noting the relatively low amount transferred by southern senders (60 tokens) when the potential partners are all from South.

Denoting with t_i the amount that sender i gives to her partner and with g_i what sender i gets back, we define the return ratio as $r_i = \frac{g_i}{3t_i}$. The third row of Table 3.2 reports the average return ratios chosen by the northern (r_N) and southern (r_S) receivers to which a sender i makes a transfer, for each value f_i^N.[5] On average, the return ratio chosen by northern receivers is just 2 percentage points higher than the one chosen by southern receivers (56% vs. 54%). Finally, we define the overall payoff earned by sender i as $\pi_{i,send} = 100 - t_i + g_i$, while the payoff for the same subject viewed as a receiver is defined as $\pi_{i,rec}$ = (sum of total amounts received by other players − sum of total amounts returned to other players). Thus the total payoff for subject i is the sum $\pi_{i,total} = \pi_{i,send} + \pi_{i,rec}$. The averages of these payoffs for the two regions ($\pi_{N,send}$, $\pi_{S,send}$, $\pi_{N,rec}$, $\pi_{S,rec}$, $\pi_{N,total}$, $\pi_{S,total}$) are displayed in the last rows of the table. On average, and independently of the sending, receiving or overall perspective, northern subjects walk out of the game with higher payoffs.

In the following, we will investigate the way that such differences emerged in the course of the game. We will also analyze how robust these differences are and if they are statistically significant. Because we are interested in situations in which players actually had a choice, we will exclude the cases where f_i^N is 0 or 1, i.e. the cases in which a sender sees only southern or northern partners in her group.[6]

3.3.1 Discrimination against South

The level of trust by players from region H to players of region K can be measured by two indicators: (i) the propensity by which a region H player contacts a region K player to make a positive transfer and (ii) the amount of transfer made by region H players to region K players. Since the frequency of players from the two regions is not the same in each group of players, contact opportunities between regions are not uniformly distributed. Thus one has to be careful in analyzing senders' behavior in terms of both (i) and (ii).

5 Note that these figures are not available for North (South) in the cases in which only South (North) are seen by sender i.

6 Note that this exclusion does not raise concerns because group composition is random.

However, note that the group composition is determined by the computer in a completely random fashion. We can, therefore, exploit the exogenous variability of group composition to test whether the region of potential partners affects the choices of players or, on the contrary, players choose their partners independently of regional considerations.

Let R_i^N be a dummy variable taking value 1 if the receiver chosen by sender i is from North and 0 otherwise. If sender i chooses her partner disregarding the region to which the receiver belongs the following equality must hold

$$E(R_i^N) = f_i^N \tag{3.1}$$

where E denotes the expectation operator. This equality says that, if the choice is random with respect to region, on average the fraction of northern receivers chosen by a sender must be equal to the fraction of northern players seen by the sender in her group. Figure 3.1 plots the sample counterpart of the expectation on the left hand side of equation (3.1) for each value of f_i^N between .25 and .75.[7] This is done for senders in the two regions separately as well as for all senders. A point above the diagonal indicates a preference for North since it means that the average frequency of choosing North is greater than the proportion of North seen by the sender.

We find that almost all the points lie above the 45 degree line. While those corresponding to southern senders are closer to it, the points for northern senders lie much further away. Of course, in the case of all senders, the points are situated in between those of North and South. Thus the figure suggests the existence of a generalized preference for choosing a northern partner as a receiver, a preference which is stronger for northern senders. Note that since R_i^N is dichotomous, and specifically bounded from above at 1, the distance from the 45 degree line has to decrease with f_i^N even in the presence of a propensity to favor North.

In order to assess whether the deviations from random choice displayed in Figure 3.1 are statistically significant we proceed as follows. Consider the regression

$$R_i^N - f_i^N = D + u_i \tag{3.2}$$

where D is a constant term and u_i is a zero mean random noise component. Note also that f_i^N is randomly assigned. Given equation (3.1), a test for the hypothesis that senders choose recipients disregarding nationality can be framed as a test for the null hypothesis that

$$H_0 : \; D = 0 \tag{3.3}$$

which implies that $R_i^N - f_i^N$ is zero mean noise.

7 For $f_i^N = 0$, $E(R_i^N)$ is of course equal to 0. For $f_i^N = 1$, conditioning on positive transfers $E(R_i^N)$ is equal to 1.

Figure 3.1: Deviation from randomness towards North (first 4 treatments)

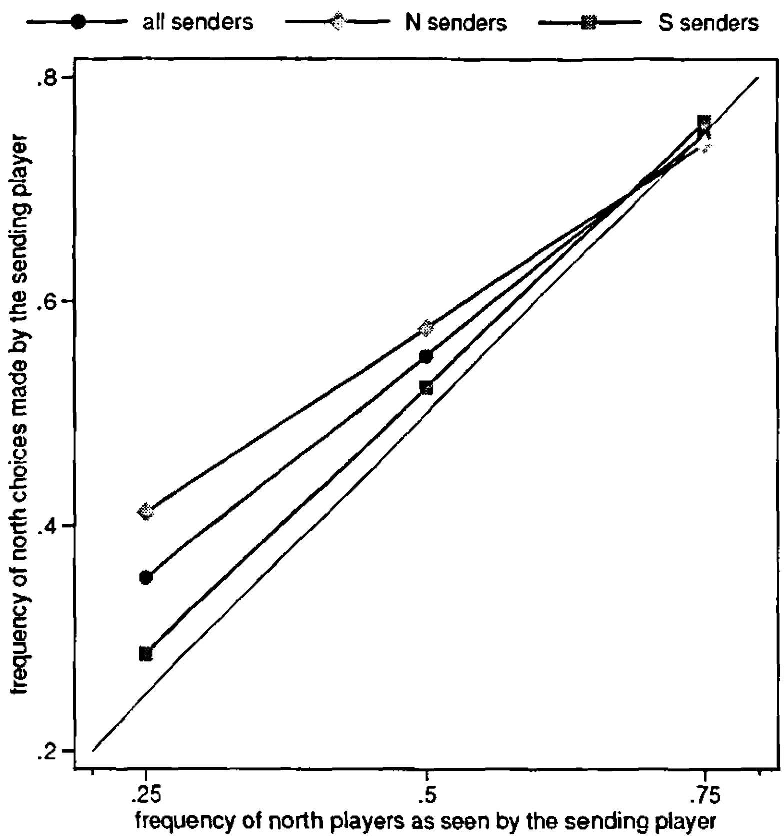


Table 3.3 reports the results of this test for the three lines displayed in Figure 3.1. The evidence of a preference towards North in choosing a partner is statistically significant when aggregating over all senders and even more so when confining only to northern senders. However, the preference of southern senders towards North is not statistically significant. Since we have repeated observations for the same sender in different periods and treatments, the standard errors are corrected to account for within-individual correlation of the error component.

Figure 3.2 displays the estimates of the constant term D in equation 3.2 for each of the first four four treatments, with 90% confidence intervals. The figure suggests that the extent of deviation from randomness in favour of North increases during the development of the game. D is not distinguishable from 0 in the first treatment, but increases in the subsequent treatments, becoming significantly different from 0 (at the 10% level) in the fourth treatment. Figure 3.3 displays the same statistic for northern senders and the tendency towards increasing discrimination against south appears even stronger.

Table 3.3: Deviations from random choice over all periods of the first 4 treatments

	All senders	N senders	S senders
D	0.044	0.062	0.022
s.e.	0.018	0.026	0.025
p-value	0.008	0.008	0.193
obs.	2167	1215	952

Note: The table reports results from the estimation of the regression $R_i^N - f_i^N = D + u_i$ by sender group. R_i^N is a dummy variable taking value 1 if the receiver chosen by sender i is from North and 0 otherwise. f_i^N is the frequency of northern players seen by sender i in her group. D is a constant parameter to be estimated. u_i is an error component. Standard errors are robust and take care of within-individual correlation of the error component. p-values are for the test that $D = 0$. The cases in which f_n^i is 0 or 1 are excluded. Note that f_n^i is randomly assigned.

No such evidence appears instead to characterize the behavior of South, as described by Figure 3.4.

Thus, the combination of results from Table 3.3 and Figures 3.2, 3.3 and 3.4 suggests the possibility that discrimination by North against South does not decrease with experience and actually builds up rather than being a strong prejudice with which North enter the game. This conjecture is further explored in the analysis that follows.

Figure 3.2: Coefficient D over treatments, all Senders

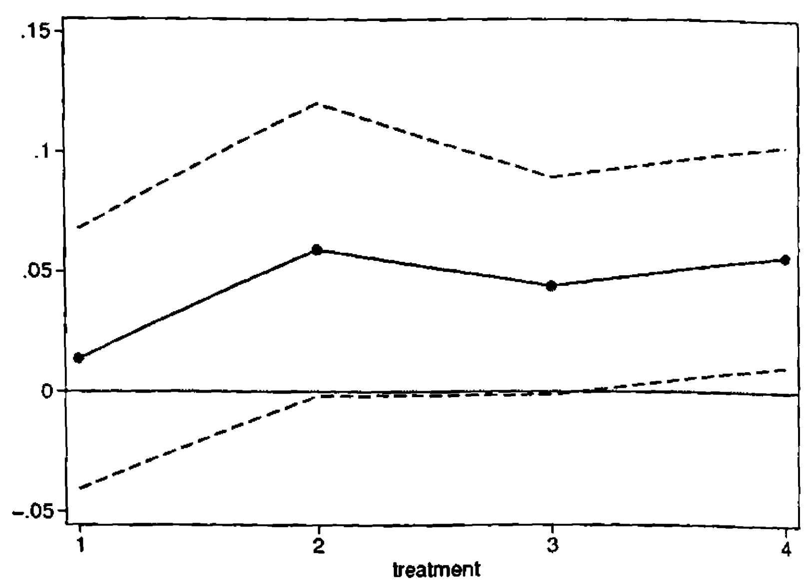


Figure 3.3: Coefficient D over treatments, Northern Senders

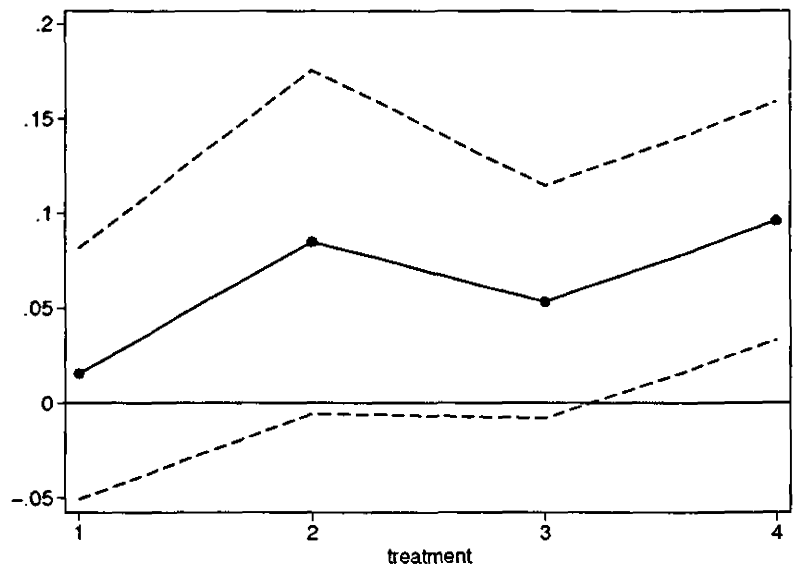
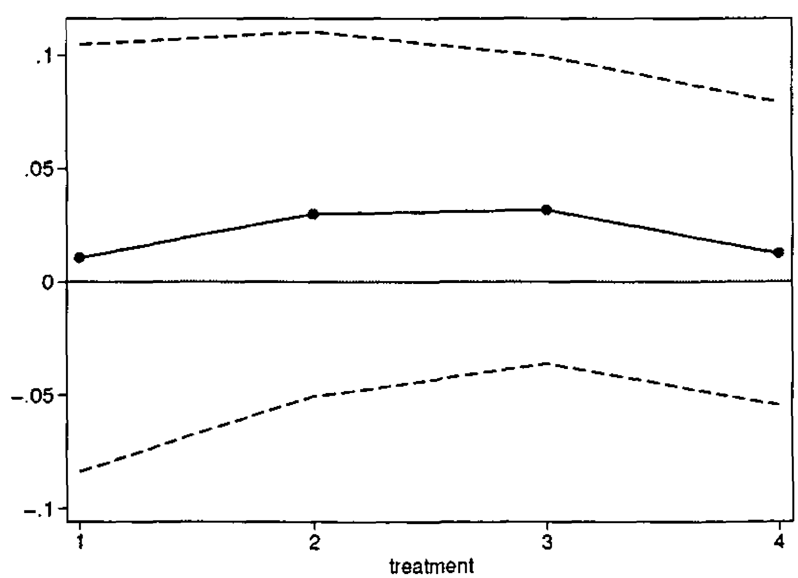


Figure 3.4: Coefficient D over treatments, Southern Senders



Next we concern ourselves with the magnitude of transfers made by senders differentiated by region. Table 3.4 shows the matrix of transfers sent by region H to region K, with $K, H \in \{N, S\}$. The top panel refers to the first four treatments, while the bottom panel refers to the earliest interaction in the game, occurring in period $p = 1$ of treatment $T = 1$. In this table, the comparison between columns within the same row indicates how the transfers received by North and by South differ. Overall, in treatments 1 to 4, North receive more than South from northern senders. However, early in the game ($p = 1, T = 1$) the opposite happens: South receive more than North from northern senders.

Table 3.4: The matrix of transfers between regions

		Southern receiver	Northern receiver	average
All periods	Southern sender	73.83	73.87	73.86
$T = 1,2,3,4$	Northern sender	81.37	86.38	84.39
	average	77.80	81.20	79.76
$p = 1, T = 1$	Southern sender	28.92	47.45	40.15
	Northern sender	63.38	54.91	57.67
	average	47.93	52.09	50.62

Note: The table reports simple averages of positive transfers from "rows" to "columns". The cases in which the fraction f_n^i of northern players seen by sender i is 0 or 1 are excluded. Note that this fraction is randomly assigned.

In order to test the statistical significance of the differences shown above, we estimate the following regression:

$$t_j = \alpha_r + \beta_r R_j^N + \delta_r X_j + \tau_j \tag{3.4}$$

where t_j is the transfer sent to receiver j, R_j^N is a dummy variable taking value 1 if receiver j is from North, X_j is a vector of dummy variables denoting the gender of the sender and of the receiver and τ_j is an error component.[8] The coefficient β_r measures the extent to which transfers received by North differ from transfers received by South. Its estimates and standard errors are reported in Table 3.5 for all senders and separately for northern and southern senders.[9]

8 The inclusion of observed characteristics like age and number of siblings does not change our results in equation 3.4 as well as in the other estimated equations that follow. At least in the case of age this is likely to be due to the lack of sufficient variation of this

Table 3.5: Differences between the transfers sent to northern and southern receivers

		By all Senders	By N Senders	By S Senders
All periods	β_r	3.24	5.44	-0.35
$T = 1, 2, 3, 4$	s.e.	1.89	2.62	2.82
	t-value	1.71	2.07	-0.13
	obs.	2330	1271	1059
$p = 1, T = 1$	β_r	4.00	-3.51	11.31
	s.e.	6.91	10.08	9.93
	t-value	0.58	-0.35	1.14
	obs.	90	51	39

Note:The table reports robust standard errors and corresponding t-values for the test that $\beta_r = 0$ in the regression $t_j = \alpha_r + \beta_r R_j^N + \delta_r X_j + \tau_j$ by sender groups. t_j is the transfer sent to receiver j. R_j^N is a dummy variable taking value 1 if receiver j is from North. X_j is a vector of dummy variables denoting the gender of the sender and of the receiver. τ_j is an error component. The estimated coefficients β_r are not numerically identical to the corresponding differences between columns of Table 3.4 because of the inclusion of controls for gender. The cases in which the fraction f_n^i of northern players seen by sender i is 0 or 1 are excluded. Note that this fraction is randomly assigned.

The first row of the table reports results for the first four treatments, taking into account the within-individual correlation of error components. It shows that, on average and controlling for gender, a northern receiver is given 3.24 tokens more than a southern one and that most of this bias is attributed to northern senders: in the second column the point estimate is 5.44. These differences are small in size but statistically significant at the 10% and 5% levels respectively. However, at early stages of the interaction (see the second row of the table) the picture is different. There is no statistically significant evidence of preferences of one group over the other and, if anything, the point estimates of β_r for the transfers sent by northern senders even indicate that North made higher transfers to South than to North.

We conclude this section by summarizing its main observations:

- South is contacted less often and receives less transfers than North, with most of this discrimination attributed to the sending behavior of North.
- The bulk of the discrimination is in the fact that South is contacted less often.

variable in our sample of young Ph.D. students.

9 The estimated coefficients β_r are not numerically identical to the corresponding differences between columns of Table 3.4 because of the inclusion of controls for gender.

- The discrimination against South does not decrease with experience and is actually less significant at earlier stages of the game compared to when it is judged based on the overall behavior.

In the next section we attempt to investigate the source of the observed discrimination and explain how it emerges. To this end we will compare South and North in terms of their payback behavior as well as their overall tendency to make transfers.

3.3.2 Why is South Discriminated Against?

We will not attempt to give a conclusive answer to this question. However, further analysis of payback behavior and the evolution of sending behavior may offer some hints. We start with three conceivable conjectures for the source of discrimination:

(A) Discrimination by North against South is a result of pure prejudice that cannot be supported by the behavior of South.

(B) Discrimination by North against South is a consequence of the fact that the return ratio of South is smaller than that of North.

(C) South receive less transfers than North because South themselves transfer little (to both North and South).

We start by comparing North and South in terms of the transfers they make to others. Going back to Table 3.4, if we compare different rows of the matrix within the same column, we see how transfers sent by North and South differ. In all cases, i.e. independently of the region of the receiver, North transfer considerably more than South. In other words North trust more all receivers. This is true when we average over the first four treatments, as well as when we look at period 1 of treatment 1. It is interesting to observe that in this early interaction North transfer more to South than to North, and South transfer very little to themselves.

To test the significance of these differences the appropriate regression to be estimated is

$$t_i = \alpha_s + \beta_s S_i^N + \delta_s X_i + \theta_i \tag{3.5}$$

where t_i is the transfer sent by sender i, S_i^N is a dummy variable taking value 1 if sender i is from North, X_i denotes the gender of the sender[10] and θ_i is an

10 The gender of the receiver cannot be included in this case because of the presence of zero transfers, i.e. situations in which there is no receiver.

error component.[11] The coefficient β_s measures the extent to which transfers sent by North differ from transfers sent by South. Results are reported in Table 3.6 for transfers sent to all receivers and separately to northern and southern receivers.[12]

Table 3.6: Differences between the transfers sent by northern and southern senders

		To all Receivers	To N Receivers	To S Receivers
All periods	β_s	10.35	12.24	5.92
$T = 1, 2, 3, 4$	s.e.	3.88	3.68	4.05
	t-value	2.67	3.33	1.46
	obs.	2310	1249	918
$p = 1$, $T = 1$	β_s	12.80	7.64	28.41
	s.e.	6.97	8.62	11.62
	t-value	1.84	0.89	2.45
	obs.	87	53	29

Note: The table reports robust standard errors and corresponding t-values for the test that $\beta_s = 0$ in the regression $t_i = \alpha_s + \beta_s S_i^N + \delta_s X_i + \theta_i$, by receiver groups. t_i is the transfer sent by sender i. S_i^N is a dummy variable taking value 1 if sender i is from North. X_i denotes the gender of the sender. The gender of the receiver cannot be included in this case because of the presence of zero transfers, i.e. situations in which there is no receiver. θ_i is an error component. The estimated coefficients β_s are not numerically identical to the corresponding differences between the rows of Table 3.4 because of the inclusion of controls for gender. The cases in which the fraction f_n^i of northern players seen by sender i is 0 or 1 are excluded. Since this exclusion restriction operates differently from the perspective of senders and receivers, the number of observations in Tables 3.5 and 3.6 differ. Note that this fraction is randomly assigned.

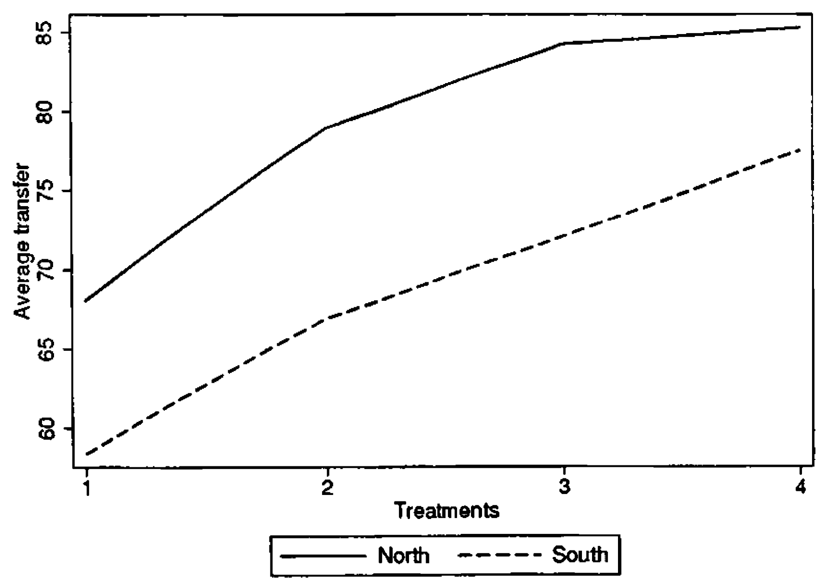
The first row of the table reveals that, controlling for gender, North transfer significantly more than South (on average 10.35 more tokens to the group as a whole, and 12.24 more tokens to northern players). North's tendency to transfer more than South is very high also at the earliest stage of the interaction (12.80 tokens more than South to the group as a whole in row $T = 1, p = 1$.) But perhaps the most interesting observation here is the fact that North treat South (in terms of transfers) much better than South treat itself (on average 28.41 more tokens in $T = 1, p = 1$).

11 Standard errors are computed taking into account within-individual correlation of the error terms.

12 The estimated coefficients β_r are not numerically identical to the corresponding differences between rows of Table 3.4 because of the inclusion of controls for gender.

In Figure 3.5 we consider the evolution of the game by looking separately at each of the first four treatments, and the evidence confirms again that North tend to transfer more than South on average.[13] If we interpret a generous transfer by a sender as an indication that the sender trusts the receiver to reward him/her later in the game (either by making a generous payback or by making a generous transfer in a subsequent period) then Table 3.6 and Figure 3.5 clearly indicate that North are endowed with the propensity to trust others more than South and that this holds also at a very early stage of the game.

Figure 3.5: Transfer comparison N and S



We next move to compare North and South in terms of their payback behavior. Here we estimate the regression

$$r_i = \delta + \gamma_r R_i^N + \eta_r X_i + \rho_i \tag{3.6}$$

where $r_i = \frac{y_i}{3t_i}$ is the return ratio chosen by the receiver for sender i. R_i^N is a dummy variable taking value 1 if the receiver chosen by sender i is from North, X_i is a vector of dummy variables denoting the gender of the sender and of the receiver, and ρ_i is an error component. The coefficient γ_r measures the extent to which northern receivers choose a higher return ratio than southern ones.

13 Note that the differences in transfers between North and South reported in the figure are statistically significant at the 5 percent level in each treatment.

Results are reported in Table 3.7 for all senders and separately for northern and southern senders.[14]

Table 3.7: Differences in the return ratio chosen by northern and southern receivers

		For all senders	For N senders	For S senders
All periods	γ_r	0.017	0.028	-0.011
$T = 1, 2, 3, 4$	s.e.	0.023	0.033	0.032
	t-value	0.75	0.86	-0.34
	obs.	2167	1215	952
$p = 1, T = 1$	γ_r	0.108	0.166	0.045
	s.e.	0.063	0.086	0.110
	t-value	1.72	1.93	0.41
	obs.	82	49	33

Note: The table reports robust standard errors and corresponding t-values for the test that $\gamma_r = 0$ in the regression $r_i = \delta + \gamma_r R_i^N + \eta_r X_i + \rho_i$ by sender groups. r_i is the return ratio chosen by the receiver for the sender i. R_i^N is a dummy variable taking value 1 if the receiver chosen by sender i is from North. X_i is a vector of dummy variables denoting the gender of the sender and of the receiver. ρ_i is an error component. The cases in which the fraction f_n^i of northern players seen by sender i is 0 or 1 are excluded. Note that this fraction is randomly assigned.

Judged on the basis of the first four treatments (see the first row of the table), the return ratio chosen by northern receivers is not significantly higher than the one of southern receivers, independently of the region of the sender. The picture changes, however, when we look at the earliest stage of the game (the first period of the first treatment) in the second row of the table. Here we see that southern receivers return significantly less than northern receivers as a fraction of what they received in the initial transfer. Moreover, note that the difference is particularly large when the sender is from North.

Taken together, the evidence provided above is against conjecture (A). We see little evidence for discrimination against South at the outset. On the other hand, South return less than North as a ratio of what they receive in the earliest stage of the game.[15] Moreover, South send lower transfers throughout the game. This suggests that the discrimination against South that builds up

14 Standard errors are computed taking into account within-individual correlation of the error terms.

15 Note that this conclusion differs from that of Fershtman and Gneezy (2001) who establish that the discrimination against Sephardi receivers is irrational as payback behavior in the two groups was essentially the same.

later in the game may have to do with South's tendency to return less at the earliest stage of the game (conjecture B) and/or to make lower transfers in general (conjecture C).

We further explore the validity of these conjectures by looking at the way contacts are positively reinforced and reciprocated from period to period. In Table 3.8 we test whether the choice of a sender in the second period differs from randomness conditioning on the choice made in the first period. We therefore estimate equation (3.2) again but this time only on the observations for the second period of all treatments and separately for the cases in which North or South were chosen in the first period. We then test the null hypothesis (3.3) for this case. Note that this hypothesis, if accepted, would imply that senders choose their partner randomly in the second period, and in particular independently of what they did in the first one.

Table 3.8: Reinforcement of deviations from random choice with respect to previous choices

period 2		all senders	N senders	S senders
N chosen in $p = 1$	D	0.169	0.212	0.101
	s.e.	0.032	0.038	0.055
	p-value	0.000	0.000	0.034
	obs.	204	125	79
S chosen in $p = 1$	- D	0.086	0.065	0.109
	s.e.	0.038	0.059	0.048
	p-value	0.012	0.136	0.013
	obs.	165	85	80

Note: This table tests whether the choice of a sender in the second period differs from randomness conditioning on the choice made in the first period. The equation $R_i^N - f_i^N = D + u_i$ is estimated, by sender groups, only on the observations for the second period of all treatments and separately for the cases in which North or South was chosen in the first period.
R_i^N is a dummy variable taking value 1 if the receiver chosen by sender i is from North and 0 otherwise. f_i^N is the frequency of northern players seen by sender i in her group. D is a constant parameter to be estimated. u_i is an error component Standard errors are robust and take care of within-individual correlation of the error component. p-values are for the test that $D = 0$. The cases in which f_n^i is 0 or 1 are excluded. Note that f_n^i is randomly assigned
The coefficients reported in "S chosen in $p - 1$" were multiplied with (-1) so that a positive sign indicates reinforcement of a southern choice.

This conclusion is, however, rejected by the evidence of Table 3.8. The first row of the table shows that if North are chosen in the first period the preference for North is reinforced by all senders in the second period. Moreover North reinforcement of a previous northern choice is greater than that of South. The evidence of reinforcement of a previous southern choice is instead very weak (see the second row of the table), particularly when the sender is from North. We conclude that North's higher standards in terms of return ratios at the outset of the game generate a stimulus that leads subjects (in particular northern senders) to reinforce a previous transfer to North. It is indeed easy to see why the initial higher standards of North in terms of return ratios should stimulate transfers to North later in the interaction. In our version of the trust game it is clearly optimal to make high transfers to subjects who have previously proven to be trustworthy. Thus, taken in isolation, the evidence of Table 3.8 would support conjecture (B).

Table 3.9: Return ratios over treatments

	$T=1$	$T=2$	$T=3$	$T=4$
$\bar{r}_N$	0.51	0.51	0.50	0.56
obs.	57	58	59	59
$\bar{r}_S$	0.47	0.49	0.48	0.51
obs.	47	47	47	50
p-value	0.46	0.78	0.60	0.22

Note: The table reports, for each treatment, the average return ratios chosen by northern and southern receivers. The reported figures are computed by first averaging the return ratios of each subject in a given treatment and then by averaging across subjects in each region, that is: $\bar{r}_{K,T=\tau} = 1/M \sum_{i=1}^{M} \bar{r}_{i,T=\tau}$, where $\bar{r}_{i,T=\tau}$ is the average ratio that subject i returned in treatment $T=\tau$ and M is the number of individuals from region K that received a transfer at least once in that treatment. K stands for North and South, respectively. p-values are for the test on the equality of $\bar{r}_N$ and $\bar{r}_S$, controlling for gender.

However, Table 3.9 confirms, from a different perspective, what was already suggested by Table 3.7: there is no indication that the initial difference in trustworthiness between North and South persists during the evolution of the game, both in terms of economic dimension and statistical significance. This table reports, for every treatment, the average return ratio chosen by each northern and southern subject for all the transfers they received.[16]

16 These figures are computed by first averaging the return ratios of each subject in a given treatment and then by averaging across subjects in each region, that is: $\bar{r}_{K,T=\tau} =$

It is important to realize that by considering the average return ratio chosen by each subject, we do not give more weight to subjects who, because of their relatively higher trustworthiness, are contacted more often. All subjects, independently of their region, are more trustworthy in the fourth treatment than in the first, and in no treatment the difference between the two regions is statistically significant. Moreover, additional computations show that in both regions the subjects who are initially less trustworthy (those who return less than the median in treatments 1 and 2) increase their return ratios equally, on average, in treatments 3 and 4 (from 0.34 to 0.41).

We know instead that throughout the entire game North transfer significantly more than South (see Table 3.6 and Figure 3.5). Thus, the difference in the propensity to trust, more than the difference in trustworthiness, appears to be likely to explain why South is discriminated against in a fashion that does not fade away during the evolution of the game. In other words, also conjecture (B) finds less support in our data. But, is it possible that the persistently higher tendency of North to trust other subjects is reciprocated by their partners with more frequent transfers to North?

Table 3.10 shows that subjects tend to trust those who trusted them. In this table we estimate how the odds that a subject i trusts a subject j depend on the existence of previous interactions between i and j. These estimates are based on the Conditional Logit Model as described extensively on page 20 in chapter 1. The first row of column 1 indicates that the odds that i transfers to j in period 2 are 3.38 times higher if i has chosen j already in period 1, as opposed to choosing another partner. The second row of the same column shows that the odds of the same event are even higher (6.83) if j has transferred to i in period 1 as opposed to making another choice. Both estimates are highly statistically significant. Thus, the subjects of our study are more likely to transfer not only to partners that they have previously trusted, but also to partners who trusted them.

The second column of Table 3.10 interacts the dummies of the first column with an indicator for whether the transfers or the return ratios received from j in period 1 were larger than the corresponding median levels in the sample. The estimated odds ratios for these interactions are larger than 1 and highly significant as well. The first one (3.32) indicates that if i chose j in period 1, i is more likely to go back to j in period 2 if j previously returned more than the median in the sample. Thus, as expected, higher trustworthiness is rewarded with a higher likelihood of a transfer in subsequent periods.

$1/M \sum_{i=1}^{M} \bar{r}_{i,T=\tau}$, where $\bar{r}_{i,T=\tau}$ is the average ratio that subject i returned in treatment $T = \tau$ and M is the number of individuals from region K that received a transfer at least once in that treatment. K stands for North and South, respectively.

Table 3.10: Reciprocation of previous positive experiences

	period 2		all periods	
	1	2	3	4
$d_{ijp-1} = 1$	3.38 (.39)***	2.02 (.31)***	3.34 (.16)***	1.74 (.12)***
$d_{jip-1} = 1$	6.83 (1.06)***	4.56 (.85)***	4.51 (.30)***	2.93 (.27)***
$d_{ijp-1} = 1 \wedge 1\{r_{ijp-1} \geq \text{med}(r)\}$	.	3.32 (.79)***	.	3.71 (.39)***
$d_{jip-1} = 1 \wedge 1\{t_{jip-1} \geq \text{med}(t)\}$	.	3.06 (.82)***	.	2.38 (.27)***
Obs.	1728	1728	8180	8180
Pseudo R^2	0.26	0.3	0.26	0.3
Correct Predictions	0.63	0.64	0.60	0.62

Note: $d_{ijp-1} = 1$ denotes the event that i has chosen j in the previous period, and $d_{jip-1} = 1$ means j has chosen i in the previous period. "$\wedge$" is the logical "and" operator, and $1\{...\}$ is the indicator function which takes value one if the expression inside the parenthesis is true. "med()" is the median of the variable. Reported values are odds ratios, standard errors in parenthesis. *, ** , *** denote significance equal to the 10, 5 and 1 percent level of a test that the odds ratio is one. Control variables included are: age and siblings of all 4 players, gender and nationality of i interacted with the attributes of j. Pseudo R^2 is the percent of variance explained by the model compared to a model which includes a constant only. Correct predictions indicates the share of observations in which the highest estimated probability $\hat{p}_j$ coincides with the actual choice.

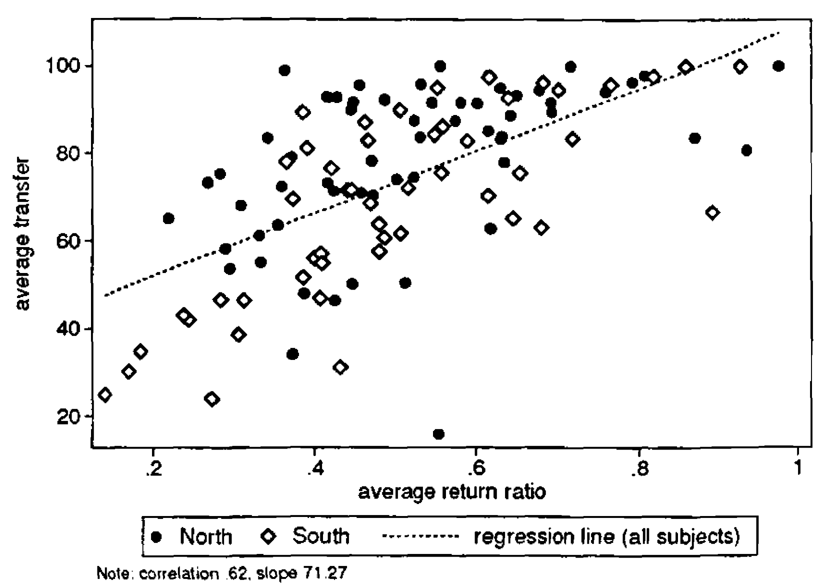
The second interaction is, however, more interesting for our purposes. It indicates that if j transferred to i in period 1, i is more likely to transfer to j in period 2 if the previous transfer from j was larger than the median in the sample.

Thus, also a higher degree of trust from others is rewarded with trust. The same analysis is replicated in columns 3 and 4 for all periods with similar results.

Interestingly trust is rewarded with trust independently of the region of the subject who trusted first. Results not reported for brevity indicate that when South and North contact another player in period 1, they are equally more likely to be contacted by this player in period 2. On the other hand northerners have a significantly higher tendency to reciprocate a contact from any other player. So the extent to which trust is rewarded with trust depend on the region of the subject who is called to reciprocate. More specifically, trust is rewarded with trust by all subjects, but more so by North than by South.

We have seen the tendency of all subjects, and in particular of northern subjects, to reward with a generous transfer those from whom they previously received high transfers. There is more than one reason for doing so. This behavior can be a consequence of a natural desire to return a favor by a favor, or it may be a tool to signal others that a high level of transfers is expected in the future as well. But there is also a third possible explanation: Figure 3.6 shows that "trust" and "trustworthiness", as measured respectively by average transfers and average return ratios are highly correlated in the subjects of our study. Hence it is possible that subjects are reciprocating with high transfers to those from whom they received high transfers in the past because they are aware of this statistical correlation and realize that the return ratios from these individuals can be expected to be high. It is reasonable to assume that all these three explanations play a role in motivating subjects to make high transfers to those who made high transfers to them.

Figure 3.6: Correlation of Trust and Trustworthiness, all Senders



Summing up, our evidence suggests that the higher northern tendency to trust all other subjects throughout the game stimulates a higher frequency of transfers to North from all other subjects and in particular from northerners. In other words, it suggests that even if southern and northern standards in terms of return ratios are very similar, the fact that South fail to approach

North's standards in terms of transfers sustains the reluctancy of North to contact South with high transfers throughout the game. We therefore conclude that conjecture (C) receives the highest support in our data. Put differently, more than for not being trustworthy, South are being punished for their own low level of trust.

Our finding concerning South's level of trust is consistent with Knack and Keefer (1997) who seek to find the correlation between social capital and trust on the one hand and economic performance on the other. Their analysis builds on the World Values Surveys that contain questionnaire data on thousands of respondents from 21 countries. To assess the level of trust they rely on the following question posed in each country: "Generally speaking would you say that most people can be trusted, or that you can't be too careful in dealing with people" Their measure of trust is defined to be the percentage of respondents replying "most people can be trusted". All but two countries (Poland and Greece) represented in our experiment appear in the survey. Excluding these two countries, the average trust measure for North is 45.4% and only 26.8% for South.

Similar conclusions are also reached by Guiso et al. (2003), who find that disparities in relative trust between people of different countries affect the level of trade. While they explain their evidence just as a consequence of stereotyping, our results indicate that differences in trust may emerge and be reinforced by repeated interactions between nationalities, even when agents are not characterized by strong stereotyping at the outset of the interaction.

Moreover note that while both Knack and Keefer (1997) and Guiso et al. (2003) base their findings on questionnaires, our findings are based on revealed preferences that emerge from subjects' decision making.

3.4 Payoffs

On average, making a transfer pays off well. Even when disregarding the fact that a subject increases his/her chance of being made a high transfer in a subsequent period by making a high transfer, subjects' payback behavior generates positive profits. On average, subjects made 60 Cents profit on every Euro transfer in payback only. South are therefore being punished for their low level of trust. Figure 3.7 plots the average payoffs (in points) of North and South at each of the six periods within treatments 1 to 4. Figure 3.8 shows the average payoff across periods for each treatment. These two figures reveal that North dominate South in terms of payoffs at each and every period (averaged over treatments) and at each and every treatment (averaged over periods).

Figure 3.7: Payoff comparison N and S. per period

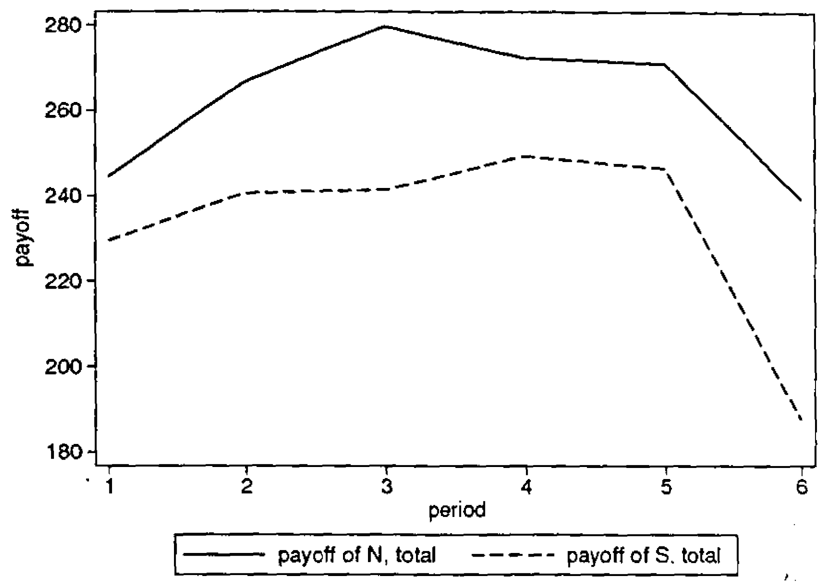
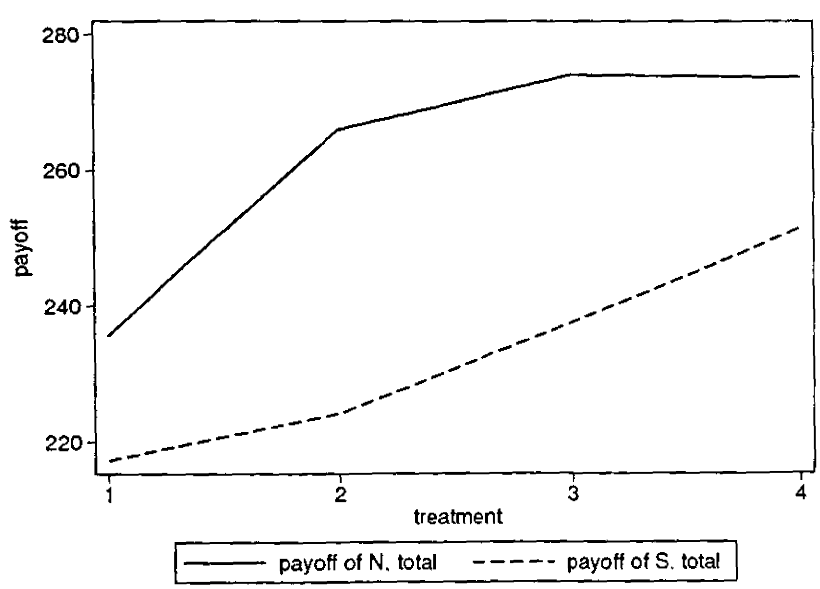


Figure 3.8: Payoff comparison N and S. per treatment



3.5 The consequences of forcing interactions

It is interesting to observe that forcing interactions appears to reduce the differences between the two regions. This is suggested by Table 3.11 where we compare descriptive statistics for the treatments $T = 4$, $T = 5$ and $T = 6$. As explained in Section 3.2, the fourth treatment is the last one of the initial series of treatments in which subjects had free choice of partner. In treatment $T = 5$ they were instead matched randomly with another subject, while in the last treatment they had again free choice. Not surprisingly, the impossibility to choose the partner reduces considerably the degree of trust and trustworthiness and therefore the average payoffs. But the most interesting result of this table is that after being forced to interact without choice of partner, subjects from the two regions appear to behave more similarly than they did before. If we compare the first and the last columns of this table, we see that the differences between the two regions in terms of transfers and payoffs are considerably smaller in treatment $T = 6$ than in treatment $T = 4$, and this happens even if average transfers go back to the levels observed before treatment $T = 5$.

Table 3.11: Descriptive statistics for treatments $T = 4$, $T = 5$ and $T = 6$

	$T = 4$	$T = 5$	$T = 6$
t_N	85	68	82
t_S	78	65	80
r_N	0.56	0.38	0.48
r_S	0.51	0.39	0.50
$\pi_{N,total}$	274	226	264
$\pi_{S,total}$	251	241	259

Note: t_K is the average transfer made by senders of region K, r_K is the average return ratio chosen by receivers of region K, and $\pi_{K,total}$ is the average total payoff earned by subjects of region K. K is equal to N or S denoting North and South respectively. In treatment 5 senders were not allowed to choose a partner.

3.6 Discussion

We have discovered significant differences between southern and northern Europeans in a dynamic version of the trust game. South is discriminated against, mainly by North, as overall it is contacted less often and ends up leaving the experiment with lower payoffs. We suggested that the observed inferior treatment that South receive can find its roots in South's own behavior in the game. South pay back less on transfers it receives at early periods of the game and make substantially lower transfers than North practically throughout the game.

While South have a slight bias in favor of North, this bias is not statistically significant as is the case for North's bias in favor North. This difference between North and South can be explained by the principle that "Losses loom greater than gains" which follows from Kahneman and Tversky (1979)'s prospect theory: We have seen that a higher reciprocity standard prevails in the North (where by reciprocity we include both the return ratios and the propensity of making high transfers in the future.) It is reasonable to assume that these differences in reciprocity also reflect different expectation about reciprocity in each region. This means that on average when North make a transfer to South, North are disappointed by the outcome (they make a loss with respect to their expectations), whereas transfers from South to North leave South with gains (relative to their expectations). Because losses loom greater than gains the forces that drive North away from South are greater than those which drive South away from South, which explains why the discrimination against South appears stronger for Northern players.

It would be a serious challenge to provide an encompassing explanation of the different standards of North and South in terms of both trust and trustworthiness as emerged from our experiment and from the evidence of Knack and Keefer (1997) and Guiso et al. (2003). While this is outside the scope of this paper we suggest two directions here: The first possibility is that these differences emerge merely from an income effect: Assuming that "generosity" and "reciprocity" are luxury (normal) goods, people will tend to "consume" more of them the greater is their income. Thus the higher level of income and stage of development in the North during recent history would be responsible for cultural differences regarding trust and trustworthiness, reflected in our results.

The other possible explanation is that differences in terms of trust and trustworthiness between South and North have to do with the different role of the family in these two regions. In both social and economic activities the family plays a much greater role in the South than in the North.[17] With family ties less intensive in the North, people in the North rely on networking outside the family more than people in the South. Trust and trustworthiness outside the family is thus more crucial for social and economic success in the North.

We point out that regardless of the preferred explanation and even if both the income effect and the family effect are weak, a convergence to two substantially different population equilibria in two societies can emerge from a grain of difference that reinforces itself in a dynamic trajectory that leads to substantial differences. This suggests the possibility that a small group of individuals endowed with low trust and trustworthiness can cause a snowball effect by which more and more people adopt their standards as trust and trustworthiness pays off less and less.

Our findings have two types of implications. Firstly, the fact that agents' choice of partners in economic interactions is not arbitrary and may depend on characteristics that appear to be payoff irrelevant (region in our case) is a message worth taking into the theoretical literature. Secondly, our findings make a valid point in the European perspective. The significance of the regional role as established in our experiments highlights the question of whether the persistence of cultural and national diversity across regions within unified Europe should not impose any impediment to achieving economic and political uniformity.

17 See, for example, Bentolila and Ichino (2003) and their references.

Part II

An Assessment of Microsimulation Techniques and Panel Unit Root Tests: Applications to Schooling Decisions and Exchange Rates

CHAPTER 4

On the Use of Ex-ante Evaluation Techniques: The Case of School Enrolment in PROGRESA

4.1 Introduction

The evaluation of policies is of particular importance in the context of development economics, where programmes aiming at poverty reduction and the improvement of the human capital of the poor are being implemented at a large scale. This importance derives from the key role that is attributed to poverty reduction in the process of economic and social development[1] and from simple considerations of political accountability. Hence, policy makers need tools to assess the impact of such programmes. Ideally, evaluations should not only be carried out after the completion of a project, but also prior to its implementation. Ex-ante evaluation techniques give answers to questions regarding the direction and the magnitude of effects that are likely to occur upon implementation of a policy. These methods thus help to improve the design of the policies and to avoid policy failures. By providing quantifiable results at early stages of project planning, these tools are an important part of impact analysis and programme monitoring. Ex-ante evaluations are carried out using microsimulation methods.[2] To date, little is known about the ac-

1 See, for example, the World Development Report 2000 on "Attacking Poverty" (Worldbank, 2000).

2 Microsimulation is a general term for methods designed to simulate systems at the level of individual units rather than the overall population. Different models require different

curacy of microsimulations. The use of such techniques is gaining popularity and an assessment of the forecasting abilities seems urgently needed. However, the evaluation of evaluation methods is a nontrivial task. This is mostly because very rarely situations occur in which the actual impact of a policy can be compared to the predictions of a microsimulation, simply because it is difficult to identify the impact of the policy in the first place. The reasons is that often a counterfactual to which the microsimulation could be compared to is not readily available.

This paper tries to fill this gap, and, for a particular type of programme, an evaluation of a microsimulation technique is carried out. More specifically, a microsimulation of a conditional cash transfer programme is done using the data from PROGRESA. Because this programme was implemented as a randomized experiment, the actual effect of the policy is easily identifiable and it thus offers a perfect benchmark for the evaluation.[3] PROGRESA is a well known programme in rural Mexico, which is in place since 1997.[4] It aims at improving the educational attainment and health status of the poor rural population. One of the main goals of the programme is to increase secondary school enrolment ratios. To this end, eligible families are offered a cash transfer conditional on the school attendance of their children. The rationale is that low school enrolment ratios are mainly due to the necessity of children to contribute to household income in poor families.

In Latin America, the daily activities of 17 percent of the children aged between 5 and 14 (some 16 million children) are considered as child labour (ILO, 2004). There is a clear distinction between between child work and child labour. While child work comprises light activities that are not considered harmful to the educational opportunities of the child, all activities that damage children's physical and psychological health and are detrimental to its future development, in particular education, are defined as child labour (Anker, 2000). Hence, the above figures point to an alarming problem in developing countries. Lack of education substantially narrows future employment opportunities and thus makes it less likely for children to break the poverty – no education circle. It is here where the effect of a conditional cash transfer programme begins. School attendance, especially at secondary level, comes at a cost for the household. Children that go to school cannot contribute to household income – be it through domestic activities or through

microsimulation techniques.

3 In that sense, the approach taken in this paper is similar to the study by Lalonde (1986) who also uses experimental data as a benchmark to evaluate the performance of the difference–in–differences estimator for differently constructed counterfactuals.

4 It is now called OPORTUNIDADES. To give some idea of its magnitude, the yearly budget equals roughly 0.2 percent of Mexican GDP (Attanasio et al., 2001).

formal labour income.[5] In addition, even if education itself is free, school attendance is often associated with additional costs for transportation and school materials, such as books. In this trade off between future and current income poor households that would decide against education for their children are offered a cash incentive to send their children to school.

In this paper the likely effects of such a programme are simulated using data from a survey conducted before the introduction of the programme. A model of occupational choice is used to simulate the potential impact of PROGRESA on school enrolment ratios. To avoid any suspicion of working backwards from the data to the method, an existing modelling framework (Bourguignon et al., 2002) will be used and no new method will be proposed. Moreover, the microsimulation exercise will follow closely the aforementioned paper. In this model, children decide between three occupational choices: they either work, work and attend school, or go to school only. The contribution of children to the income of a household through domestic work is explicitly accounted for. The model identifies the key elements that are necessary to simulate the impact of the programme.

This paper shows that the predictions of the model come close to the real effect. Moreover, disaggregated by age and gender, the real effect is within bootstrapped confidence intervals of the simulation exercise. While the overall performance of the microsimulation is good, some critical aspects of the method become apparent. In particular, the uneven distribution of the sample over the three occupational choices causes the model to have a poor fit.

The remainder of the paper is organized as follows. The next section introduces the model and discusses how it can be used for an ex-ante simulation. Section 4.3 presents the results of the simulation exercise. In Section 4.4 results from a standard ex-post evaluation are presented and related to similar findings in the literature. Ex-ante and ex-post results are compared and some lessons from the simulation are drawn in Section 4.5 before the last section concludes.

4.2 Ex-ante Evaluation: Theory

Choosing a model for a microsimulation exercise involves finding a compromise between structural sophistication and feasibility. While some structure is necessary in order to identify the effect of relevant variables and relate them to the variables through which the programme operates, one has to bear in mind

5 Especially in in poor rural areas children often have to contribute to household production, either by working on the fields or by taking care of younger siblings, and for these reasons are not sent to school (OECD, 2003).

that the ultimate goal is to simulate a programme, which requires tractability and robustness across specifications. However, structural models have the advantage of offering further insight into individual behaviour through the identification of key parameters. On the other hand, many aspects of a fully structural model, while interesting, are often irrelevant for the simulation, as long as the structure of the model is flexible enough to accommodate them. In addition, many parameters made explicit in structural approaches often cannot be identified due to the simple lack of convincing instruments, which further reduces the empirical value of rich structural models.

The background from which theories of child labour and schooling decisions may draw ranges from household time allocation decisions to lifetime income considerations and the effects of human capital accumulation through (foregone) education. Attanasio et al. (2001), for example, develop a structural model in which each child maximizes its lifetime earnings and takes its schooling decisions accordingly, given a market wage and some intertemporal discount factor. The conditional transfer affects the rate at which children substitute their earning losses against education. In this context, the authors study the impact of varying levels of an educational grant, given some value for the discount factor. While appealing from a modelling point of view, one could argue that in rural Mexico the trade-off around which their model is built is often dominated by liquidity constraints. Poor households in rural Mexico do not have access to the – poorly developed – financial intermediaries that could facilitate this intertemporal decision. This, in combination with the lack of lifetime planning horizons in poor families, can be seen as a limitation of the applicability of their model.

Todd and Wolpin (2003) develop a dynamic structural model at the household level. This approach addresses the major shortcoming of other models that optimize at the individual level by considering interactions between decisions taken within a family. In their model, parents maximize parental lifetime income by choosing an optimal fertility rate and deciding upon the time allocation of their children. Fewer but better educated children are seen as an alternative that becomes more attractive the lower the costs of education. The framework hence addresses more fundamental questions than the mere schooling decision and could be useful for assessing the long-term impact of the cash transfer. In the present case, where the time horizon is of two years, it is however unlikely to observe alterations in the fertility behaviour.

Finally, Bourguignon, Ferreira and Leite (2002) develop a model of a discrete labour supply decision, where a child, contributing to the household income, decides either to go to school, to work, or to mix between the two options. The trichotomous framework is especially suitable for developing countries where many children do both, attend school and are engaged in some kind of economic activity, especially in rural areas (see for example OECD (2003), Table A11). The reduced form of their model also encompasses contributions of children to home production, an important but often neglected aspect that is prevalent among the rural poor. The model relates household income and children's earnings to the labour supply outcome for each child. It is silent with respect to the other aforementioned aspects such as interactions at the household level or lifetime earnings considerations. The model was developed around the *Bolsa Escola* programme in Brazil, a conditional cash transfer scheme which shares the main characteristics of PROGRESA.

The aim of this paper is to shed some light on the performance of microsimulation techniques. Because the outcome of the simulation will be compared with the already known effect of the programme, it does not seem appropriate to propose an entirely new framework for the ex-ante evaluation of PROGRESA. Any new suggestion would inevitably be prone to criticism for being the outcome of a process that moved from the data to the theory rather than *vice versa*. Hence, after the options outlined above, the general framework used in this paper is taken from Bourguignon et al. (2002). Moreover, in a first step, the method developed in their paper will be closely followed to ensure a high degree of comparability.

4.2.1 A Model of Occupational Choice

The model of Bourguignon et al. (2002) (henceforth BFL. See also Bourguignon et al. (2003)) is based on the following assumptions. The unit of decision is the child and not the household. Hence, considering the intra-household labour allocation, it is assumed that schooling decisions of children are taken independently from each other, and labour market outcomes of other family members are unaffected. Further, the composition of households is taken as exogenous, and effects on fertility are disregarded.

Let the variable S_i be a choice variable that takes one of the following values j for each child i, $i = 1, ...N$:

$$S_i = \begin{cases} 0 & \text{if } i \text{ does not attend school} \\ 1 & \text{if } i \text{ works and attends school} \\ 2 & \text{if } i \text{ attends school and does not work.} \end{cases}$$

Under choice $j = 0$ it is assumed that individual i works full time, either being engaged in domestic activities or by working on the market. Similarly, under choice $j = 2$ it is possible that children allocate some time to domestic work in a way that will be described below.

In the framework of a multinomial choice model, each i will make an optimal choice according to:

$$S_i = k \text{ if } S_k(.) > S_j(.) \text{ for all } j \neq k,$$

where $S_j(.)$ is some function $S_j(X_i, H_i, Y_{ij}, v_{ij})$ of the following variables:

X_i : characteristics of i (age, schooling *etc.*)
H_i : characteristics of the household of i (parental education *etc.*)
Y_{ij} : income of the household of i when choosing j
v_{ij} : i.i.d. shock.

The household's income Y_{ij} is the sum of i's own income y_{ij} and the income of the rest of the household, Y_{-i}. After combining all non-income variables into $Z_i = [X_i, H_i]$ and linearizing the model it takes the following random utility representation:

$$U_{ij} = \gamma_j Z_i + \alpha_j(Y_{-i} + y_{ij}) + v_{ij}. \tag{4.1}$$

Even though this specification does not explicitly model all aspects surrounding the schooling decision, it contains the relevant variables through which the programme operates, namely the household income (which will be affected by the transfer) and the income of a child working on the market. Note also that the coefficients γ_j and α_j may vary with each choice.

Assume for the moment that the potential earnings of i and hence its potential contribution to the household income is observable for each child and denote this amount by w_i. Then, depending on the amount of time dedicated to work, the actual contribution y_{ij} can be written according to the choices j:

$$\begin{aligned} y_{i0} &= K w_i \\ y_{i1} &= M y_{i0} = MK w_i \\ y_{i2} &= D y_{i0} = DK w_i. \end{aligned}$$

These expressions indicate that under choice $j = 0$ a fraction K of these potential earnings is realized. Under choice $j = 1$, where the child attends school and works, only a fraction MK of w_i is realized. If a child goes to school ($j = 2$), it may contribute to domestic production for a value of DK times the potential market earnings. Combining the above with equation (4.1) yields

$$\begin{aligned} U_{i0} &= \gamma_0 Z_i + \alpha_0 Y_{-i} + \beta_0 w_{i0} + v_{i0} \\ U_{i1} &= \gamma_1 Z_i + \alpha_1 Y_{-i} + \beta_1 w_{i1} + v_{i1} \\ U_{i2} &= \gamma_2 Z_i + \alpha_2 Y_{-i} + \beta_2 w_{i2} + v_{i2} \end{aligned} \tag{4.2}$$

with $\beta_0 = \alpha_0 K, \beta_1 = \alpha_1 MK, \beta_2 = \alpha_2 DK.$

If potential earnings could be observed for each i and estimates of the coefficients α_j were available. the model in equation (4.2) could be used for microsimulation by looking at the effect of an exogenous variation of the household income under the schooling options $j = 1, 2$.

4.2.2 Estimation and Identification

In reality, however, earnings are only observed for those children who work for remuneration. While domestic work may contribute substantially to the household's production, it remains unobservable. But the representation in equation (4.2) requires some value for w_i for each child, no matter if the income is realized on the market or through home production. To overcome this the observed market earnings under choices $S_i = \{0, 1\}$ will be used to impute potential earnings for those children that either work at home or go to school. Following the standard wage equation literature (Mincer, 1974), the observed earnings w_i can be explained by:

$$\log w_i = \delta X_i + m \cdot \mathbf{1}\{S_i = 1\} + u_i. \tag{4.3}$$

The vector X_i contains the standard regressors for a wage equation, in particular age, years of schooling, *etc.* and δ is the corresponding parameter vector. The indicator function $\mathbf{1}\{...\}$ accounts for the fact that earnings of a child under option $S_i = 1$ might be significantly lower because some time m is spent at school. In practice, a control for sample selection (*e.g.* (Heckman, 1979)) is necessary to control for possible biases. For those children whose wages are not observed, a potential wage $\tilde{w}_i$ can be imputed using the estimates of equation (4.3) and drawing a random element of the residuals vector $\hat{u}_i$. This gives a complete description of the earnings vector w.

Assuming exponentially distributed errors, the choice model in equation (4.2) is known as the multinomial logit model (McFadden, 1973). In this model, the coefficients are identified only relative to a certain choice category. In the following, choice $j = 0$ was chosen as the base category. In this case, the model yields estimates of the relative coefficients $(\gamma_k - \gamma_0)$, $(\alpha_k - \alpha_0)$ and $(\beta_k - \beta_0)$, $(k = 1, 2)$. In general, this is not a problem. But since the cash transfer is state dependent, it is necessary to identify all three coefficients $(\alpha_0, \alpha_1, \alpha_2)$ that are related to the household income in order to do a microsimulation. Following an argument from BFL, the coefficients can be identified by making one simple structural assumption. Call the estimated coefficients from the multinominal logit model $\dot{a}_1$ and $\dot{a}_2$. Then it follows that

$$\begin{aligned}
\alpha_1 - \alpha_0 &= \dot{a}_1 \\
\alpha_2 - \alpha_0 &= \dot{a}_2 \\
\alpha_1 MK - \alpha_0 K &= \dot{b}_1 \\
\alpha_2 DK - \alpha_0 K &= \dot{b}_2
\end{aligned}$$

Notice that in equation (4.3) the coefficient of the indicator function gives an estimate of M, namely $\dot{M} = \exp^{\dot{m}}$, because this is an estimate of the difference in earnings realized under alternative $j = 1$. With an estimate of M, any arbitrary combination of K and D allows to identify the coefficients α_j as:

$$\begin{aligned}
\dot{\alpha}_0 &= \dot{\alpha}_1 - \dot{a}_1 \\
\dot{\alpha}_1 &= \frac{\dot{a}_1 - \dot{b}_1/K}{1 - \dot{M}} \\
\dot{\alpha}_2 &= \dot{\alpha}_1 + \dot{a}_2 - \dot{a}_1 \\
\dot{D} &= \frac{\dot{b}_2/K + \dot{\alpha}_0}{\dot{\alpha}_2}.
\end{aligned}$$

Assume that $K = 1$, *i.e.* children who do not attend school realize their full potential earnings either in the market or through home production. The estimates $\dot{a}_1$ and $\dot{a}_2$ can easily be transformed into the structural parameters α_j of the model.

4.2.3 Impact Simulation

Having identified the levels of the income coefficients α_j, the impact of the cash transfer is simulated using the following conditional payment:

$$\begin{aligned}
U_{i0} &= \gamma_0 Z_i + \alpha_0 Y_{-i} + \beta_0 w_i + v_{i0} \\
U_{i1} &= \gamma_1 Z_i + \alpha_1 (Y_{-i} + TR_i) + \beta_1 w_i + v_{i1} \\
U_{i2} &= \gamma_2 Z_i + \alpha_2 (Y_{-i} + TR_i) + \beta_2 w_i + v_{i2},
\end{aligned}$$

where TR_i is the transfer paid conditional on school enrolment, which in turn depends on i's characteristics. At this stage it becomes clear why an identification of the α_j coefficients is necessary. Being *conditional* on choices 1 and 2, the transfer is asymmetric. The identified difference of coefficients $\alpha_j - \alpha_0$ would not not be enough to simulate the impact.

In multinomial choice models, the residual terms $v_{i1} - v_{i0}$ can neither be observed not precisely estimated. However, for each i, the set of residuals

v_{i0}, v_{i1} and v_{i2} are bound to belong to certain intervals, such that given the parameter estimates and i's characteristics they are consistent with the actual choice. In particular, if choice $j = 0$ is taken as a base choice (this means its utility is normalized to 0) and individual i has chosen choice $j = 1$, it must be that:

$$U_{i1} - U_{10} > 0 \text{ and } U_{i1} - U_{12} > 0,$$

which can be expressed as

$$\begin{gathered} Z_i(\gamma_1 - \gamma_0) + Y_{-i}(\alpha_1 - \alpha_0) + y_{ij}(\beta_1 - \beta_0) + (v_{i1} - v_{i0}) > \\ \mathrm{SUP}[0, Z_i(\gamma_2 - \gamma_0) + Y_{-i}(\alpha_2 - \alpha_0) + y_{ij}(\beta_2 - \beta_0) + (v_{i2} - v_{i0})], \end{gathered}$$

where SUP[ ] denotes the supremum of the expression. This places a condition on each pair of errors $\{(v_{i1} - v_{i0}), (v_{i2} - v_{i0})\}$, each of which is drawn from a double exponential distribution. This condition ensures that residuals are choice consistent respecting the error distribution used in the multinomial logit specification.[6]

4.3 Ex-ante Evaluation of PROGRESA

The available data on PROGRESA was collected in five waves between November 1997 and November 1999.[7] The first two waves of data available are the pre-programme survey data and were collected in November 1997 and in March 1998. However, only the first wave asked income related questions and questions regarding the occupational choice. Hence, only this wave will be used to perform the ex-ante simulation. The focus group are children between 10 and 16 years old. As documented elsewhere in the literature (Schultz, 2001), the school enrolment of the 6 to 9 year old is almost 100 percent.

In the first wave 17.5 percent of the respondents between 10 and 16 years old reported to receive payment for some work. Table 4.1 summarizes the weekly wage by age. To put these numbers into perspective, the household income per capita is also reported. It shows that children who work make a significant contribution to the household's monetary income, and this contribution is increasing with age. In comparison, in Brazil BFL find similar figures, although the relative contribution there is lower. One has to bear in mind that households in rural Mexico tend to be rather large and include many (inactive) members, which affect the per capita figures.

6 See appendix B.3 on how to draw the residuals.

7 The reader who is not familiar with the main features of PROGRESA is referred to Appendix B.1. A detailed description of how the variables were derived from the survey questionnaire can be found in Appendix B.2.

Table 4.1: Average weekly earnings and per capita household income

age	obs.	average earnings	average pc income	median earnings	median pc income
10	572	33	42	27	30
11	550	38	42	27	30
12	423	52	43	27	31
13	379	70	46	39	32
14	450	111	47	100	34
15	704	149	53	120	36
16	878	152	58	140	41
10-16	3,956	95	47	60	33

Note: Table reports statistics of those children that are reported to work for remuneration. All values in Mexican pesos. 10 pesos roughly corresponded to 1.10 US Dollar in that period.

Table 4.2 gives the realizations of the choice variable that may take three values according to the occupation chosen. A breakdown by gender shows that girls are less likely to attend school throughout all age groups. On average, 26.2 percent of the boys and 31.2 percent of the girls do not attend school. Enrolment ratios drop significantly after the age of 11. What is evident is that the option to work and go to school at the same time is particularly predominant in younger age groups, highlighting the necessity to allow for a trichotomous choice.

Table 4.2: Reported status

	age							
boys	10	11	12	13	14	15	16	10-16
not school	2.5	4.4	11.1	21.7	34.3	51.3	66.8	26.2
work and school	17.2	18.4	13.6	10.2	7.6	6.3	5.6	11.5
school only	80.3	77.2	75.3	68.1	58.1	42.5	27.6	62.3
observations	1,818	1,667	1,806	1,665	1,628	1,623	1,477	11,684
girls	10	11	12	13	14	15	16	10-16
not school	3.0	4.4	17.4	30.5	44.6	60.4	73.2	31.2
work and school	16.8	17.4	10.2	6.7	3.4	2.4	2.0	8.9
school only	80.2	78.2	72.4	62.8	52.0	37.2	24.8	59.9
observations	1,724	1,700	1,611	1,581	1,490	1,493	1,293	10,892

Note: Values indicated the percentage share each cell. For a definition of status, see text.

Table B.2 in the appendix reports the sample means of some variables by occupational choice. Several interesting aspects become apparent. On average, children who do not go to school or work and go to school come from households with higher monetary income, supporting the previous evidence that their work contributes significantly to the endowment of the household. Those who attend school are, on average, younger and have a higher rank.[8] Those who do not go to school tend to come from states with a higher median wage, although there is considerable variation across gender and age not reported in the table, reflecting regional variations in child labour demand. Unsurprisingly, children who go to school tend to come from households with higher parental education.

There is substantial variation across age and gender, which, in the case of PROGRESA, where the transfer schemes differ for boys and girls, becomes even more relevant. Even though the data set is very rich with 22,576 observations, a breakdown by gender, age, and occupational status would lead to some combinations containing very few observations. The estimations reported in this section capture this difference by reporting results for various age groups. The 10-11 years old form the first group, the 12-13 years old the second and the 14-16 years old constitute the last group. Within each group, of course, gender differences are accounted for using indicator variables.

4.3.1 Estimation of the Earnings Vector

Inspection of the model in equation (4.2) reveals that the market earnings of each child, potential or realized, are an important ingredient in determining the decision to go to school or not. However, as mentioned earlier, market wages are only observed for those children that report to work on the market and have to be imputed for all other individuals. This section reports the estimates of equation (4.3).

Clearly there might be a potential sample selection bias when looking only at those children who work. The issue of sample selection becomes difficult if one considers the selection into any of the three categories of the choice model.[9] A feasible way of dealing with this is to consider just the decision to work or not to work as a potential source of a bias. In the present case, a simple Heckman procedure to correct for sample selection proves to work fine and shows that, indeed, there is a selection effect.

8 The rank is defined as the position of the child with respect to all household members below the age of 19. For example, a child with rank 3 has 2 elder siblings, and a child with rank 1 is the oldest child in the household below 19.

9 Bourguignon et al. (2001b) discuss the mechanism suggested by Lee (1983).

Results for one age group are presented in Table 4.3.[10] Note that two exclusion restrictions are included in the selection equation. The rank of the child and the presence of the father in the household are assumed to influence the decision to work or not work only, but not to influence the wage. This seems plausible as there is no reason to believe that these factors affect market earnings. Interestingly, the higher the rank of a child the lower is the probability of working. This confirms the view that first-born children are sometimes disadvantaged in this respect. Also, the absence of a father in the household makes it more likely for children to work.

Table 4.3: Estimation of the earnings equation

log(earnings)	coefficient	t-value
female	-0.42	-3.08
years of schooling	-0.07	-1.50
(years of schooling)^2	0.004	0.87
median state earnings	0.74	6.07
status (1 if S_j=1)	-0.91	-19.88
selection		
rank	-0.09	-3.23
father in hh	-0.22	-5.29
female	-0.77	-23.23
years of schooling	0.13	3.44
(years of schooling)^2	-0.02	-6.03
median state earnings	-0.17	-1.59
$\dot{\rho}$	0.62	
$\dot{\sigma}$	0.85	
$\dot{\lambda}$	0.53	2.37
Number of obs	8241	
Censored obs	6299	
Uncensored obs	1942	

Note: Table presents estimates for equation (4.3), using a Heckman two step estimation to correct for sample selection. Only children in the age group of 14-16 years. Age dummies and constants included but not reported.

10 A full set of results for all age groups is available upon request.

In the wage equation, the variables have the expected effect. Females earn significantly less than males, and, as expected, those who work and go to school also earn significantly less as compared to those that just work. One important variable that accounts for the regional variation in child labour is the median earnings of children of that particular age/gender group in the respective state. This variable, which proves to be significant, captures all age and local specific demand shifts to child labour. Additional years of schooling increase the wage, and the wage is also increasing in age. Those effects are not always significant, in particular because the regression is carried out by age group, and consequently he age induced variation is limited. In addition, the degree of schooling is very similar across the age group.

In order to impute potential wages for those that do not work, earnings were estimated by relying on the coefficients of that particular age group and adding a random draw from the estimated residuals vector from the earnings estimation.

4.3.2 Estimation of the Choice Model

Coefficients and t-values of the estimation of the multinomial choice model for one age group are given in Table 4.4. The coefficients were found to vary significantly across different ages, and it was not possible to fit a model that would cover all age groups in a satisfactory way. In this estimate, the outcome $S_j = 0$ is taken as a reference group.

Table 4.4: Estimation of multinomial logit model

	work and school		school only	
	coefficient	t-value	coefficient	t-value
Y_{-i}	0.0001	2.75	-0.0002	-0.35
w_i	-0.0045	-4.09	-0.0035	-8.82
total members of household	-0.059	-1.58	0.028	1.78
years of schooling	-0.860	-1.44	0.065	0.24
(years of schooling)^2	0.048	2.33	0.037	4.03
(age-years of schooling)^2	-0.025	-1.17	0.010	1.04
female	-1.192	-8.92	-0.475	-9.13
max parental education	0.007	0.28	0.085	7.70
children below 6 in household	0.055	0.66	-0.136	-3.82
rank	-0.232	-2.19	0.045	1.07
median state earnings	-1.272	-2.77	-1.193	-6.48

Note: Table reports estimates of the multinomial logit for the group of 14-16 year old. Outcome "not school" is the comparison group. 8236 observations, Pseudo $R^2 = 0.13$. Age dummies and constant included but not reported.

The fit of the multinomial choice model varies greatly across age groups. Inspection of Table 4.2 reveals that for the very young the proportion of the choice $S_i = 0$ (not school) is very small, whereas for the older children the option $S_i = 1$ (school and work) only has a small share. This asymmetry between the three choices is reflected in the poor fit of the model. The model's prediction is not satisfactory as can be seen from Table 4.5. This table compares for each of the three outcomes the actual with the predicted outcome, disregarding the error term. A breakdown by ages shows that in most cases the choice of the smallest categories is entirely attributable to the error term, but not the estimated coefficients.

Table 4.5: Accuracy of model prediction

	predicted outcome			
original outcome	not school	work / school	school only	total
not school	3,290	16	2,533	5,839
work and school	164	735	1,030	1,929
school only	1,389	229	11,647	13,265
total	4,843	980	15,210	21,033

Note: Table is a cross tabulation of the actual status and the predicted status. The share of correct predictions is 70 percent.

As discussed in Section 4.2.2, identification requires an assumption about K, which is the amount of time dedicated to household production for those children that do not work for a market wage. Assuming $K = 1$, one obtains estimates for each age group for the structural parameters α_j and β_j. Table 4.6 illustrates the differences across age groups. Note that the coefficients α_j, which determine the reaction of each individual to household income other than i's decrease significantly with age, falling from 0.126 to 0.011. This large difference between age groups does not depend on the choice of K. Compared to BFL who also find a decreasing effect of Y_{-i} with age, the results here are slightly more pronounced. Estimates of M range around 34 to 40 percent, in line with previous findings. All coefficients are in the plausible range with the estimates of D being very close to unity for the two younger age groups.

Table 4.6: Estimation of structural parameters assuming $K = 1$.

age group	M	D	α_0	α_1	α_2	β_0	β_1	β_2
10 - 11	0.380	0.996	0.125	0.126	0.125	0.125	0.048	0.125
12 - 13	0.335	0.918	0.026	0.027	0.026	0.026	0.009	0.024
14 - 16	0.402	0.619	0.011	0.011	0.010	0.011	0.004	0.006

4.3.3 Impact Simulation

The PROGRESA transfer TR_i is determined for each i by the gender and the years of schooling completed and transformed into a weekly subsidy.[11] The individualistic approach of this model makes it impossible to account for the maximum transfer that was payable for each household. In order not to contaminate the results, all household in which the transfer would have surpassed the monthly limit were dropped from the analysis.

Simulation of the transfer is straightforward. While the utility under choice $j = 0$ remains unchanged (the programme's conditionality), under choices $j = 1, 2$ the term $\alpha_j TR_i$ has to be added to the index. Together with the estimated residuals this gives a new utility level $\tilde{U}_j$ and an option j is chosen such as to maximize utility. The difference of the distributions under U_j and $\tilde{U}_j$ determines the impact of the programme.

Table 4.7: Estimated transition matrix

	simulated outcome			
original outcome	not school	work / school	school only	total
not school	3,576	117	992	4,685
work and school	0	1,582	1	1,583
school only	0	6	10,815	10,821
total	3,576	1,705	11,808	17,089

Note: Table presents the transition from the observed occupational choices (rows) to the simulated choices after the programme (columns).

This is done for each age group separately and according to i's characteristics. Table 4.7 gives the overall transition from the actual to the simulated status. The impact of the programme is clearly visible. The number of children not going to school decreases, and the decrease is captured mainly by an increase in the number of those that go to school only, while a few individuals switch to the option work and school. A closer look at the effect by gender and age group will be done in Section 4.5.

11 See Table B.1 in Appendix B.1 for details.

4.4 The Benchmark: Ex–post Analysis

The effect of PROGRESA on school enrolment ratios has been extensively analysed in the literature. The findings of this section are in line with the previous studies, such as Schultz (2001), Behrman et al. (2001), and Attanasio et al. (2001), and will therefore be presented in a concise way. The comparison between school attendance before and after the programme was launched will be confined to making a cross section comparison between wave 5 (November 99) and waves 1 and 2.

4.4.1 Pre–programme Differences

A common finding is that, in PROGRESA, by and large, the randomization seems to be successful in that the control and treatment group do not exhibit large differences at the outset. A closer look at pre-programme differences in school enrolment reveals that with one exception there are no significant differences. Tables B.4 and B.5 in the Appendix report the average enrolment ratios by age in the control and in the treatment group, separately for boys and for girls.

As can be seen from Table B.4, while the overall sample exhibits one significant positive difference between treatment and control group in the category of 13 years old, restricting the analysis to the eligible sample, none of the differences are significant. Indeed, the differences between control and treatment group are very small in value and range from -0.6 to 4 percentage points. In contrast, for girls (Table B.5) there is one significant difference in the pre-programme values, where the treatment group exhibits a significantly lower value (-4.5 percent) in the category of 12 years old. In the the next age category (13 years) the value of the treatment group is higher (4 percent, although not significant). Except for these two outliers all differences are negligible.

4.4.2 Difference Estimation

If the randomization of the experiment was successful, the difference-in-differences (DD) estimator should coincide with the simple difference estimator (D), which compares the treatment with the control group after the programme has been completed. In particular, since both groups were affected by the same macroeconomic shocks during the time period in question and there were no significant pre-programme differences between the groups, the simple D estimator identifies the effect.

The difference estimator (D) compares, by age and gender, the difference between the enrolment ratios of the control and the treatment group at the end of the programme.[12] Tables 4.8 and 4.9 provide an estimate of these differences. Table 4.8 shows that the enrolment ratio among eligible boys (last column) has risen, on average, by 4.9 percentage points. The increase was most pronounced in the age groups 13 to 15 years (up 8 points) and with just 2 percentage points it is lowest among the 10 to 11 years old. A look at the non-eligible population confirms that the targeting of the programme was successful, because with the exception of the 13 years old these do not differ between the two populations. A similar picture emerges for the girls where the average effect was higher with 6.1 percentage points. Among the eligible the effect was highest for the 13 to 15 years old and, as for the boys, lowest in the youngest age group, where it is also insignificant.

Table 4.8: Difference estimator, boys

age	10	11	12	13	14	15	16	10-16
				full sample				
difference	0.017	0.019	0.050	0.045	0.040	0.082	0.041	0.013
t-value	**2.52**	**2.36**	**3.69**	**2.44**	1.79	**3.18**	1.60	**5.46**
				not eligible				
difference	-0.003	0.010	0.041	-0.094	-0.041	0.065	0.031	0.009
t-value	-0.19	0.44	0.96	**-2.00**	-0.71	1.08	0.50	0.43
				eligible				
difference	0.020	0.020	0.051	0.070	0.054	0.085	0.044	0.049
t-value	**2.69**	**2.37**	**3.60**	**3.47**	**2.25**	**2.99**	1.55	**5.78**

Note: Table reports for each cell the difference between in enrolment ratios between the treatment and the control group respectively, and the corresponding t-values.

Figure 4.1 pictures the D estimator for eligible children by age and gender and compares it to the DD estimator. Inspection of the figure shows that especially for girls the two estimates are almost identical. Notice also that the kink in the DD estimate for 13 years old is a reflection of the pre-programme differences in that particular age group. For boys, the D estimate is slightly above the DD estimate but does follow the same pattern.

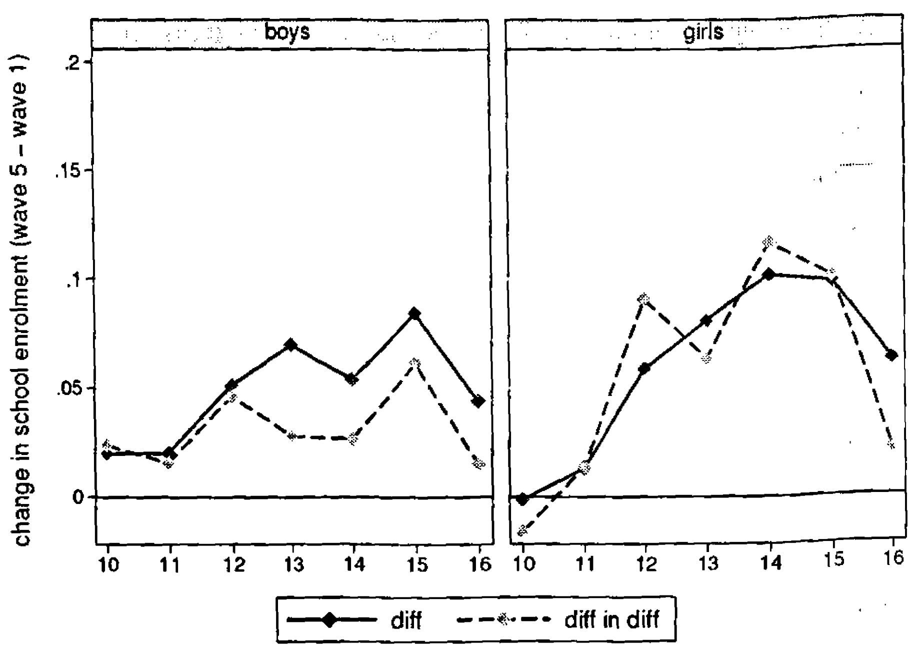
12 In the present case this is wave 5.

Table 4.9: Difference estimator, girls

age	10	11	12	13	14	15	16	10-16
				full sample				
difference	0.001	0.006	0.061	0.084	0.077	0.069	0.058	0.053
t-value	0.14	0.62	**3.90**	**4.16**	**3.20**	**2.64**	**2.11**	**6.26**
				not eligible				
difference	0.015	-0.032	0.079	0.105	-0.042	-0.048	0.041	0.008
t-value	0.75	-1.38	1.81	**1.93**	-0.74	-0.81	0.65	0.37
				eligible				
difference	-0.001	0.013	0.059	0.081	0.102	0.098	0.062	0.061
t-value	-0.21	1.15	**3.48**	**3.70**	**3.84**	**3.37**	**2.04**	**6.69**

Note: Table reports for each cell the difference between in enrolment ratios between the treatment and the control group respectively, and the corresponding t-values.

Figure 4.1: Effect of PROGRESA: D vs. DD estimate



4.5 Comparison of the Results and Discussion

This section brings together the results from the ex-ante and the ex-post evaluation. In order to do so, the trichotomous choice variable is reduced by the work dimension so that it only reflects whether a child attends school or not. Figure 4.2 visualizes the results of the simulation and compares it to the difference estimate, by gender and age. The corresponding numbers can be found in Table 4.10. In addition, the figure depicts 90 percent confidence intervals around the simulations, which were obtained using the bootstrap mechanism described in Section B.4.

The first observation is that the simulated and the real effect are more or less of the same magnitude and follow a similar pattern across ages. The simulation peaks for the 13 year old children, both for boys and for girls. The next observation is that the simulated effect is above the difference estimate for young ones and switches below the real effect for the 15 and 16 years old. Overall, while the difference estimate proposes an increase in school enrolment ratios by 5.4 percentage points, the simulated effect is 6.4 points. Hence, in general the simulation can be regarded as a good approximation of the real effect.

Figure 4.2: Simulated effect and D estimate

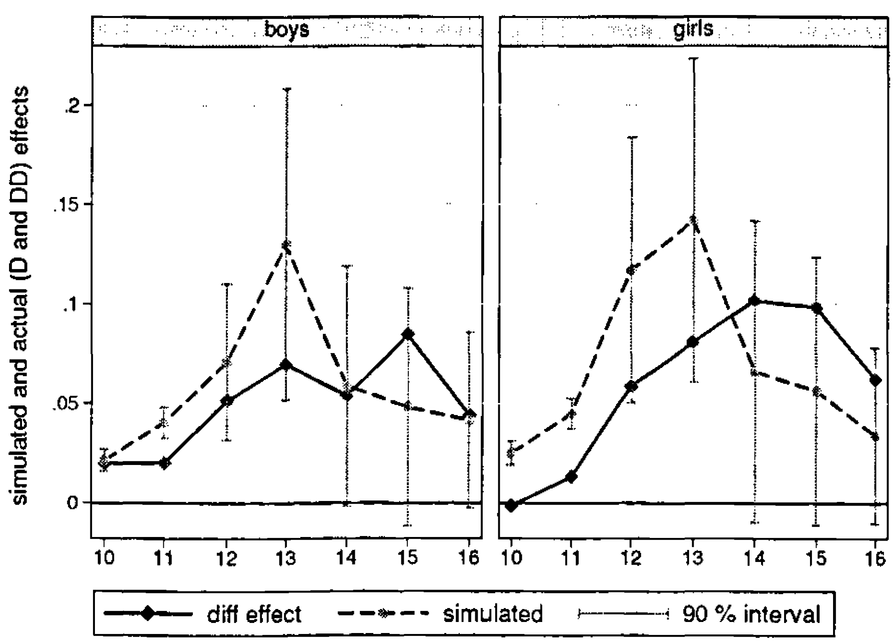


Table 4.10: Simulation, D and DD estimates

	age	10	11	12	13	14	15	16	10-16
boys	before	.975	.952	.885	.789	.645	.476	.361	.726
	sim. impact	.021	.041	.071	.130	.059	.048	.042	.059
	D impact	.020	.020	.051	.070	.054	.085	.044	.049
	DD impact	.024	.016	.046	.028	.026	.062	.015	.031
girls	before	.967	.953	.815	.686	.538	.389	.271	.660
	sim. impact	.025	.045	.117	.142	.066	.056	.034	.069
	D impact	-.001	.013	.059	.081	.102	.098	.062	.059
	DD impact	-.016	.014	.091	.063	.116	.101	.021	.056
total	before	.971	.953	.850	.738	.591	.432	.316	.693
	sim. impact	.023	.043	.094	.136	.062	.052	.038	.064
	D impact	.009	.017	.055	.075	.078	.091	.053	.054
	DD impact	.004	.015	.068	.046	.071	.081	.018	.043

Note: Tables reports the enrolment ratios in the respective situation of the eligible household.

The reason why for some age groups the predictions deviate from the actual effect are related to the relatively poor fit of the multinomial choice model. As can be seen from Table 4.2, the three outcomes of the dependent variable are very unequally distributed for each age group. For example, among the 10 and 11 years old, school attendance is clearly dominating and the fraction of children that work but do not go to school is very small. The same is true for the category "work and go to school" among the 15 and 16 years old. This, in turn, causes the estimates of the multinomial choice model to have poor predictive power for these categories. In fact, for the young ones, the estimated model would not predict any of the individuals to be in the "not school" group. Those that do not go school are driven into that category by large individual shocks. This has consequences for the simulation exercise where many more children are presumed to go to school than actually do.

If three categories are distributed unevenly, maybe a dichotomous framework would yield better predictions. The obvious variable to focus on in this context would be school attendance. In the case of the 10 and 11 years old this would of course not solve the problem of a very small non attendance group, because enrolment ratios are almost 100 percent from the beginning, but it might be useful to do so for the elder children. This requires a new setup of the simulation framework.

One important point where the procedure in this paper departs from BFL is the issue of sample selection problems when estimating the earnings equation (4.3). In contrast to the findings of BFL, for the case of PROGRESA this analysis finds that a simple Heckman correction mechanism proves to work well. Indeed, there is a significant selection effect which influences the estimated results. Not accounting for it would lead to inconsistent estimates of the coefficients, which, in turn, would give biased estimates of the potential earnings for those individuals that do not report earnings.

Inspection of the transition matrix in Table 4.7 shows that the programme is also likely to reduce the incidence of child labour. Some 90 percent of the children that will start going to school are predicted not to work on the market, whereas 10 percent will do both, go to school and work. The contribution to household income - be it through monetary income or home production - will be much lower if children attend school, hence, reducing the time they spend working.

4.6 Concluding Remark

This paper applies a microsimulation method to evaluate the impact of a conditional cash transfer program and compares the effect with the actual outcome of the policy. The simulation correctly predicts that school enrolment ratios among the target population will increase as a result of a cash payment for school attendance. Hence, the model is a valuable tool to asses the effects of such schemes which, by increasing educational attainment, are also likely to reduce poverty in the long run.

A discrete model of occupational choice is used which allows for three outcomes: attend school, work, or do both. The model picks up the main mechanism through which the transfer affects the schooling decision, considering explicitly actual and potential market earnings of each child. While the overall performance of the microsimulation is good, a closer inspection reveals that the method does not perform well if the population is unequally distributed over the three categories. This is a common problem of choice models. One solution is to employ a microsimulation method for a dichotomous framework. While this comes at the cost of loosing important features of the estimates such as the effect of the programme on the amount worked, it might improve the forecasting abilities with respect to school enrolment ratios.

CHAPTER 5

On the Use of Panel Unit Root Test on Cross-Sectionally Dependent Data: An Application to PPP

5.1 Introduction

Panel unit root tests are becoming a standard tool in the analysis of mostly macroeconomic panels. Two procedures, the Levin, Lin and Chu (2002)[1] and the Im, Pesaran and Shin (1997) test for unit roots are among the most popular. The tests have been applied to a range of macroeconomic problems, *e.g.* to the question whether real exchange rates are random walk processes or not (*e.g.* O'Connell (1998), Papell (1997)) or to investigate the mean reversion properties of the current account (Wu, 2000). Evans and Karras (1996) use panel unit root tests to analyze the convergence of regions in the US using a modified Levin et al. (2002) test procedure, while Strauss (2000) addresses the question of permanent components in regional GDP using these panel unit root tests.

However, relatively little is known about the size and power properties of these tests when any of the distributional assumptions underlying their construction is violated. The asymptotic distribution of both test statistics relies on the independence of the sections of the panel. This assumption might often be violated in real data, especially in a macroeconomic context.

1 A previous version of this test was known as Levin and Lin (1993). See also Levin and Lin (1992).

Given their widespread use, it is important to know more about the reliability of the test results. The impact of such dependence on the performance of the tests is studied in this paper. Two different forms of sectional dependence are considered. In the short run (Section 5.3.1), positive cross-sectional dependence of the error terms is analyzed. It is found that in the case of *common* shocks, eliminating common time effects is remedy enough to restore the size properties reasonably well. In fact, the test statistic does converge to a standard normal distribution. In this respect this paper contrasts the finding of O'Connell (1998), who attests severe size distortions to the Levin et al. (2002) test in the presence of common contemporaneous correlation. When the contemporaneous correlation takes different forms, however, severe size distortions do occur. Long-run sectional dependence might be present if the series of the panel are cointegrated (Section 5.3.2). In this case, the series are nonstationary but share a common stochastic trend. Early work on the study of this effect on panel unit root tests has been done by Crowder (1997) in a simple cointegration framework. The effect cointegration has on unit root test is analytically studied in Lyhagen (2000). However, data generating process considered in this paper resembles more the one considered in Banerjee et al. (2000). In line with the results of these studies, it is found that the tests are oversized as a consequence of cointegration, as long as the errors are kept independent. In Section 5.3.3 cointegration is combined with sectional correlation, *i.e.* long- and short-run dependence are brought together. This seems natural as there is no prior reason to believe that these phenomena should be mutually exclusive. The result is surprising. Considered separately, long- and short-run dependencies tend to yield oversized test results. If brought together, under some parameter configurations the size distortions go in the opposite direction: the over–rejection of the null hypothesis of a unit root vanishes and the tests become undersized. As a result, without further knowledge about the data generating process, panel unit root tests in presence of sectional dependence are inconclusive.

The application in Section 5.4 contributes to the purchasing power parity debate by addressing the question of mean reversion in a panel of real exchange rates. A panel of 18 exchange rates is first analyzed by estimating the contemporaneous covariance matrix of the error terms and corresponding standard errors. Different ways of estimating covariance matrices in the presence of heteroscedasticity and serial correlation are discussed and a bootstrap algorithm developed by Politis and Romano (1994) is suggested as a way of obtaining standard errors for these estimates. To know whether long-run sectional dependence is present in the data, a cointegration analysis following Johansen (1995) on a subset of exchange rates is conducted. Together with the simulation results obtained earlier, the existence of both dependencies in the data puts a big caveat on the use of panel unit root tests in this context in particular, and on cross-sectionally dependent data in general.

5.2 Panel Unit Root Tests

The test developed by Levin et al. (2002) (henceforth LLC) can be seen as a natural extension of the Dickey and Fuller (1981) test for a unit root to a set of time series. It builds on the method previously suggested by Quah (1990) and Breitung and Meyer (1991). In the light of the criticism by Pesaran and Smith (1995) of the use of pooled regressions of the LLC type, Im et al. (1997) (henceforth IPS) allow for heterogeneity of the series under the alternative and do not make use of traditional panel estimation techniques. They propose instead a group-mean Lagrange multiplier test and a group mean t-test based on the individual ADF test statistics. The asymptotic properties for both tests are derived by assuming a diagonal path limit. The behaviour of the cross-section dimension (N) and the time dimension (T) are functionally tied, *i.e.* $(T(N), N \to \infty)$. For LLC, as both go to infinity, T increases faster than N, such that $N/T \to 0$, whereas IPS only require $\sqrt{N}/T \to 0$.

This section presents the framework for the analysis of panel unit root tests. As in the univariate case, three forms of deterministics are considered starting from the following data generating process (DGP) that yields nonstationary series if the autoregressive coefficient ρ_i is equal to one:

$$\Delta x_{it} = (\rho_i - 1)x_{it-1} + \mu_i + \beta_i \cdot t + \epsilon_{it}.$$

The index i indicates the section of the panel ($i = 1, ..., N$) and the time index t ranges from 1 to T. The constant of each section is denoted by μ_i and $\beta_i \cdot t$ represents a time trend in the data. The assumptions on the error term ϵ_{it} are discussed further below. Table 5.1 summarizes the *a priori* restrictions and the hypotheses to be tested in each of the three models. The most general specification, model $m = 3$ in the classification of LLC, is designed to discriminate between a set of I(1) processes with drift under the null and a set of trendstationary processes under the alternative. In model 2, the trend parameter is restricted to zero *a priori*. It is used to discriminate between a set of I(1) processes without drift under the null and allows stationary processes with an expected value different from zero under the alternative. This model will be used throughout the Monte Carlo study. In the simplest model, under the null hypothesis of a unit root, x_{it} is a set of

Table 5.1: Different models and hypotheses

model	*a priori*	Null-hypothesis and alternative
$m=3$		$H_0^{(3)}: \rho_i = 1\ \forall\, i\ (\Rightarrow \mu_i \neq 0, \beta_i = 0)$
		$H_1^{(3)}: \mid \rho_i \mid < 1\ \forall\, i\ (\Rightarrow \beta_i \neq 0)$
$m=2$	$\beta_i = 0$	$H_0^{(2)}: \rho_i = 1\, \forall\, i\ (\Rightarrow \mu_i = 0)$
		$H_1^{(2)}: \mid \rho_i \mid < 0\, \forall\, i\ (\Rightarrow \mu_i \neq 0)$
$m=1$	$\beta_i = 0$	$H_0^{(1)}: \rho_i = 0\ \forall\, i$
	$\mu_i = 0$	$H_1^{(1)}: \mid \rho_i \mid < 0\ \forall\, i$

I(1) processes without drift, while under the alternative it is a set of stationary processes all with an expected value of zero.

5.2.1 Levin, Lin and Chu (2002)

The LLC test is implemented in four steps.
Step 1: Elimination of time specific effects. The cross-section average at t is subtracted from the data, i.e. $x_{it} = \tilde{x}_{it} - \frac{1}{N}\Sigma_{i=1}^{N}\tilde{x}_{it}$, which is equivalent to the introduction of time specific dummy variables. This step will play a crucial role in the simulation exercise in Section 5.3.
Step 2: Computation of ADF-statistics and normalized residuals. The choice of the lags L_i to be included should be based on a common information criterion (e.g. Akaike or Schwartz) and done *after* the elimination of time specific effects. Instead of the usual equation:

$$\Delta x_{it} = \delta_i x_{it-1} + \sum_{j=1}^{L_i} \theta_{ij}\Delta x_{it-j} + \mu_i + \beta_i t + \epsilon_{it},$$

the coefficient of interest, δ_i, is estimated by partitioning the regression using the Frisch–Waugh theorem to obtain residuals from each step:

$$\Delta x_{it} = \sum_{j=1}^{L_i} \theta_{ij}^{(1)}\Delta x_{it-j} + \mu_i^{(1)} + \beta_i^{(1)} t + e_{it} \Rightarrow \hat{e}_{it}$$

$$x_{it-1} = \sum_{j=1}^{L_i} \theta_{ij}^{(2)}\Delta x_{it-j} + \mu_i^{(2)} + \beta_i^{(2)} t + v_{it-1} \Rightarrow \hat{v}_{it-1}.$$

The regression of the residuals gives an estimator for δ_i:

$$\hat{e}_{it} = \delta_i \hat{v}_{it-1} + \epsilon_{it}. \tag{5.1}$$

In order to control for heterogeneity in the variances of the series, the residuals are normalized by the standard error σ_{ei} of regression (5.1), estimated by:

$$\dot{\sigma}_{ei}^2 = \frac{1}{T - L_i - 1} \sum_{t=L_i+2}^{T} \left(\dot{e}_{it} - \dot{\delta}_i \dot{v}_{it-1} \right)^2,$$

and the normalization is done as follows:

$$\tilde{e}_{it} = \frac{\dot{e}_{it}}{\dot{\sigma}_{ei}} \qquad \text{and} \qquad \tilde{v}_{it-1} = \frac{\dot{v}_{it-1}}{\dot{\sigma}_{ei}}.$$

Step 3: Computation of the long-run variance. For each series the long-run variance is computed using the first differences:

$$\dot{\sigma}_{xi}^2 = \frac{1}{T-1} \sum_{t=2}^{T} \Delta x_{it}^2 + 2 \sum_{\tau=1}^{K} w_{K\tau} \left(\frac{1}{T-1} \sum_{t=2+\tau}^{T} \Delta x_{it} \Delta x_{it-\tau} \right). \tag{5.2}$$

The choice of covariance weights ensures positive estimates of the long-run variances. LLC suggest the Bartlett weights, $w_{K\tau} = 1 - \tau/(K+1)$. The estimate is consistent if the truncation parameter K grows exponentially at a rate less than T, LLC suggest $K = 3.21 T^{1/3}$. The ratio of the estimated long-run variation and the standard deviation is computed, which under the null approaches one. For the adjustment, the average of this ratio across sections is also needed:[2]

$$\dot{s}_i = \frac{\dot{\sigma}_{xi}}{\dot{\sigma}_{ei}} \qquad \text{and} \qquad \dot{S}_N = \frac{1}{N} \sum_{i=1}^{N} \dot{s}_i.$$

Step 4: Computation of the test statistic. Under the null hypothesis the normalized residuals $\tilde{e}_{it}$ are independent of the normalized lagged residuals $\tilde{v}_{it-1}$. This is estimated using OLS:

$$\tilde{e}_{it} = \delta \tilde{v}_{it-1} + \tilde{\epsilon}_{it}. \tag{5.3}$$

Under the null hypothesis and in model 1, the regression t-statistic t_δ is asymptotically normal, but has to be adjusted in models 2 and 3, so that, in general:

$$t_\delta^* = \frac{t_\delta - N \tilde{T} \dot{S}_N \dot{\sigma}_\epsilon^{-2} \, \mathrm{SE}(\dot{\delta}) \, \mu_{m\tilde{T}}^*}{\sigma_{m\tilde{T}}^*} \overset{H_0^{(m)}}{\sim} N(0,1),$$

2 In the case of a trend the steps above should be implemented after demeaning the differenced series.

where SE($\hat{\delta}$) is the standard error of $\hat{\delta}$, $\hat{\sigma}_\epsilon$ is the standard error of the regression (5.3). $\mu^*_{m\tilde{T}}$ and $\sigma^*_{m\tilde{T}}$ are necessary adjustments for the mean and the standard deviation. These vary according to m, the model chosen and $\tilde{T}$, the average number of observations per section in the panel adjusting for lagged differences, $\tilde{T} = T - \frac{1}{N}\Sigma_{i=1}^N L_i$ (see Table 2 in LLC).

The asymptotic properties are derived in Levin and Lin (1993, Section 4).[3] In model specifications 2 and 3 the estimator $\hat{\delta}$ has a downward bias, which is due to the dynamic specification of the panel, especially for small T and N (Nickell, 1981). This makes the mean adjustments necessary. Furthermore, under the null, the variance of the estimator $\hat{\delta}$ falls at the rate $\frac{1}{NT^2}$, reflecting super-consistency. As N grows large, the variance of $\hat{\delta}$ gets smaller and smaller, which makes the variance adjustment necessary. If not adjusted, mean and variance bias would force the t-value to negative infinity in models 2 and 3. Under the alternative, x_{it} is already stationary, so Δx_{it} has asymptotically zero variation at zero frequency, meaning that each standard deviation ratio s_i as well as the average ratio $\hat{S}_N$ becomes small. In this case the mean adjustment does not influence the t-value adjustment, so that the adjusted value diverges to negative infinity. This shows the advantage of using an estimate of the long-run variance to discriminate between stationary and nonstationary processes.

5.2.2 Im, Pesaran and Shin (1997)

The IPS test extends the LLC framework by allowing for a mixture of stationary and nonstationary series under the alternative hypothesis. The test is defined for models 2 and 3, and the alternative is modified to:

$$H_1^{(IPS)} = \rho_i < 0, \forall i = 1, 2, ...N_1, \rho_i = 0, \forall i = N_1 + 1, ..., N.$$

IPS suggest a group mean lagrange multiplier (LM) test and a group mean t-test based on the individual ADF t-values. In simulations done by the authors the t-test outperforms the LM test slightly. According to the ADF lag order chosen in each section and the length T, adjustments are necessary to the mean and variance. The test statistics becomes:

$$\Psi_{\bar{t}} = \frac{\sqrt{N}\left\{\bar{t}_{N,T} - \frac{1}{N}\sum_{i=1}^N E[t_{i,T}(L_i, 0) \mid \rho_i = 0]\right\}}{\sqrt{\frac{1}{N}\sum_{i=1}^N Var[t_{i,T}(L_i, 0) \mid \rho_i = 0]}} \overset{H_0^{(IPS)}}{\sim} N(0, 1).$$

3 See also page 139 in the Appendix for a detailed treatment of the asymptotic properties in the case $m = 1$.

The adjustments $E[...]$ and $Var[...]$ are tabulated in the paper. The expression $\bar{t}_{N,T} = \frac{1}{N}\sum_{i=1}^{N} t_{i,T}(L_i,\theta_i)$ is the mean of the actual ADF test statistics. IPS also suggest the inclusion of time specific effects in the regression or, alternatively, the demeaning of the panel at each t. Note, however, that in contrast to LLC, the IPS–test uses an average of t–statistics and not a single estimated t-value from the pooled series.

5.3 The Effect of Cross-sectional Dependence

The model considered in this paper is designed to discriminate between a set of I(1) series without drift and a set of AR(ρ) series with expectation different from zero. In terms of standard macroeconomic time series, this configuration refers to, for example, interest rates, exchange rates and possibly price indices. The DGP takes the following form:

$$\begin{pmatrix} \Delta \mathbf{x}_t \\ \Delta \mathbf{y}_t \end{pmatrix} = \mathbf{AB}' \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \end{pmatrix} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t. \tag{5.4}$$

Both $\mathbf{x}_t$ and $\mathbf{y}_t$ are $(N \times 1)$ vectors. $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}_t$ are $(2N \times 1)$ vectors. The vector of interest is always $\mathbf{x}_t$. The $\mathbf{y}_t$ are used to simulate potentially shared stochastic trends if desired. The matrices $\mathbf{A}$ and $\mathbf{B}$ determine the long-run relation between the variables and will be defined according to the set of experiments. For example, if

$$\mathbf{A} = \alpha \begin{pmatrix} -\mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{pmatrix}, \mathbf{B}' = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \text{ and } \boldsymbol{\mu} = \mathbf{0},$$

and $\alpha = 0$, then $\mathbf{x}_t$ will be a set of independent I(1) variables without drift. The short-run correlation is modeled through the error structure:

$$\boldsymbol{\epsilon}_t \sim N(0, \sigma^2\boldsymbol{\Omega}) \text{ and } \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & \mathbf{I} \end{pmatrix}.$$

Note that contemporaneous correlation only affects the vector $\mathbf{x}_t$, not $\mathbf{y}_t$. The innovation variance is chosen to be $\sigma^2 = 1$ throughout the paper. In general, the correlation matrix $\boldsymbol{\Sigma}$ takes the following form:[4]

$$E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t] = \underset{N\times N}{\boldsymbol{\Sigma}} = \begin{pmatrix} 1 & & & & \\ \omega_{21} & 1 & & & \\ \omega_{31} & \omega_{32} & 1 & & \\ & & & \cdots & \\ & & & & 1 \end{pmatrix}, \tag{5.5}$$

where the correlations are $|\, \omega_{ij} \,| < 1$.

4 Considering only the first N elements of $\boldsymbol{\epsilon}_t$.

5.3.1 Cross-sectional Correlation

The first set of experiments is designed to measure the impact of cross-sectional correlation.[5] The absence of error correlation ($\omega_{ij} = 0, \forall i, j$) produces the desired size properties, see Table C.1 in the Appendix. Once a common, positive sectional correlation is introduced ($\omega_{ij} = 0.7, \forall i, j$), the tests appears to be slightly oversized (Table 5.2), especially for small N. This contrasts sharply the findings of O'Connell (1998) who finds size distortions of as much as 50 percent for the 5 percent size. Such distortions can be reproduced if step one of the LLC test, *i.e.* the elimination of common time effects, is not carried out. The results for different values of ω are presented in Tables 5.3 and 5.4 for the LLC and IPS test respectively.

Table 5.2: Size properties with common shocks, common time effects elimiated

$\omega = 0.7$	LLC			IPS		
	nominal size 10%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.128	.115	.111	.116	.100	.104
$T=50$	.127	.112	.106	.117	.108	.102
$T=100$	.115	.110	.106	.110	.108	.101
	nominal size 5%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.065	.052	.059	.064	.052	.058
$T=50$	.067	.054	.049	.068	.053	.050
$T=100$	.060	.056	.055	.054	.058	.055
	nominal size 1%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.014	.009	.013	.016	.010	.015
$T=50$	.015	.011	.011	.016	.012	.013
$T=100$	.011	.012	.014	.014	.015	.011

Note: Based on 4,000 replications. The values reported are the percentage of rejections using the indicated nominal level. Ideally, real and nominal size should be equal.

The power of the LLC test and the IPS test was analyzed for the two alternatives $\rho = .9$ and $\rho = .95$, where $\rho = 1 - \alpha$. This exercise was repeated for varying covariance structures, $\omega = \{0, 0.7, 0.8, 0.9\}$, $N = 25$, $T = \{60, 100\}$.

5 All simulations were carried out in the software package STATA using the modules Bornhorst and Baum (2001a) and Bornhorst and Baum (2001b).

and $\boldsymbol{\mu} = 1$. The results of this analysis are reported in Table 5.3 for the LLC test and in Table 5.4 for the IPS test. They show that once common time effects are eliminated, the power of the tests is not severely affected by cross-sectional correlation. More interestingly, the distortions in power and size are independent of the degree of cross-sectional dependence.

Table 5.3: Size and power properties of LLC for varying ω and ρ

$N = 25$	nom. size			power					
				$\rho = .9$			$\rho = .95$		
$\omega = 0$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.09	.04	.005	.99	.98	.85	.56	.38	.13
$T = 100$	.01	.04	.002	1	1	1	.83	.68	.31
$\omega = 0.7$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.10	.04	.005	.99	.98	.84	.54	.37	.12
$T = 100$	.11	.04	.009	1	1	.99	.79	.64	.29
$\omega = 0.8$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.10	.04	.005	.98	.96	.80	.54	.36	.10
$T = 100$	.11	.05	.011	1	1	1	.80	.60	.24
$\omega = 0.9$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.10	.04	.005	.99	.97	.82	.54	.36	.12
$T = 100$	.11	.04	.008	1	1	1	.79	.64	.28

Note: Based on 2,000 replications. One minus the power is the probability that the test fails to reject the null if it is false for a given significance level.

The natural question that arises is why demeaning, or, equivalently, the inclusion of time dummies, seems to be such an effective instrument if errors are correlated in the way studied here. The expected value of the outer product of the error terms is (considering the relevant first N elements) $E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'] = \boldsymbol{\Sigma}$. The elimination of time effects can be rewritten as:

$$(\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_t)(\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_t)' = \left[\left(\mathbf{I} - \frac{\mathbf{l}\mathbf{l}'}{N}\right)\boldsymbol{\epsilon}_t\right]\left[\left(\mathbf{I} - \frac{\mathbf{l}\mathbf{l}'}{N}\right)\boldsymbol{\epsilon}_t\right]' \text{ where } \mathbf{l} = (1, \dots, 1)$$
$$\mathbf{Q}\boldsymbol{\epsilon}_t(\mathbf{Q}\boldsymbol{\epsilon}_t)' = \mathbf{Q}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\mathbf{Q}' = \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}'.$$

where $\mathbf{Q}$ is:

$$\underset{N\times N}{\mathbf{Q}} = \begin{pmatrix} 1-\frac{1}{N} & -\frac{1}{N} & \dots & -\frac{1}{N} \\ & 1-\frac{1}{N} & \dots & -\frac{1}{N} \\ & & \dots & \\ & & & 1-\frac{1}{N} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} N-1 & -1 & \dots & -1 \\ & N-1 & \dots & -1 \\ & & \dots & \\ & & & N-1 \end{pmatrix}.$$

Table 5.4: Size and power properties of IPS for varying ω and ρ

$N = 25$	nom. size			power					
				$\rho = .9$			$\rho = .95$		
$\omega = 0$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.11	.05	.013	1	1	.99	.86	.76	.46
$T = 100$	.11	.05	.010	1	1	1	1	.99	.95
$\omega = 0.7$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.12	.07	.010	1	1	.99	.87	.77	.47
$T = 100$	.12	.05	.008	1	1	1	1	.99	.93
$\omega = 0.8$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.12	.07	010	1	1	.99	.89	.78	.47
$T = 100$	.12	.05	.006	1	1	1	1	.99	.93
$\omega = 0.9$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 60$	.12	.07	.010	1	1	.99	.87	.77	.46
$T = 100$	.11	.05	.009	1	1	1	1	.99	.94

Note: See Table 5.3.

If $\mathbf{\Sigma}$ takes the form where all off-diagonal elements are equal to ω, the above expression further simplifies to:

$$\mathbf{Q\Sigma Q}' = \frac{1-\omega}{N}\begin{pmatrix} N-1 & -1 & \cdots & -1 \\ & N-1 & \cdots & -1 \\ & & \cdots & \\ & & & N-1 \end{pmatrix} =$$

$$= (1-\omega)\frac{N-1}{N}\begin{pmatrix} 1 & \frac{-1}{N-1} & \cdots & \frac{-1}{N-1} \\ & 1 & \cdots & \frac{-1}{N-1} \\ & & \cdots & \\ & & & 1 \end{pmatrix} \underset{N\to\infty}{=} (1-\omega)\mathbf{I}. \tag{5.6}$$

After demeaning, the degree of cross-sectional correlation (the value of ω) leaves the relation of the off-diagonal to the diagonal elements unchanged, but it is this relation which determines the degree to which independence is violated. It is therefore not surprising that the LLC and IPS test do not show significant differences in power and size for varying ω. Moreover, for reasonable large N, the off-diagonal entries are small, e.g. with $N = 20$ the remaining 'effective' correlation is -0.05. For large N this approaches zero, just as it is in the absence of any cross correlation. In fact, as shown in Appendix C.1, the test statistic approaches a standard normal distribution.

This argument is limited, however, to the special form of Σ where all $\omega_{ij} = \omega$. Because this might not always be the case, the correlation matrix is now chosen to be a band matrix, where the correlation coefficient decreases with the distance from the main diagonal. The idea behind this specification is that there might be some natural ordering of the sections, reflecting e.g. the geographical distribution of units in a spatial model. Errors are more correlated the closer two sections are:

$$E[\epsilon_t \epsilon_t] = \underset{N\times N}{\Sigma} = \begin{pmatrix} 1 & \omega^1 & \omega^2 & \omega^3 & \dots & \omega^{N-1} \\ & 1 & \omega^1 & \omega^2 & \dots & \omega^{N-2} \\ & & 1 & \omega^1 & \dots & \omega^{N-3} \\ & & & 1 & \dots & \\ & & & & \dots & \omega^1 \\ & & & & & 1 \end{pmatrix}. \tag{5.7}$$

Table 5.5 reports the effect of this disturbance has on the performance of the LLC test, given $\omega = 0.7$ and varying N and T. It shows that the test performs quite poorly. Increasing N seem to worsen the results.

Table 5.5: Size properties of LLC with errors as in equation (5.7)

	nom. size 10%		nom. size 5%		nom. size 1%	
$\omega^i = 0.7^i$	$N = 10$	$N = 25$	$N = 10$	$N = 25$	$N = 10$	$N = 25$
$T = 20$	.227	.235	.156	.170	.066	.080
$T = 60$	.249	.250	.170	.180	.064	.078
$T = 100$	.258	.252	.177	.180	.068	.084

Note: Based on 10,000 replications.

Short-run correlation of this type does affect the size properties, no matter if common time effects are eliminated or not. In the case of common effects, the distortions are far less worrisome than previously claimed.

5.3.2 Cross-Sectional Cointegration

There are several parameters that influence the specific form of cointegration that one can observe in a vector of time series. One aspect is the number of cointegrating vectors (CIVs) in a system, or, complementarily, the number of stochastic trends driving it. Another set of parameters are the values of the loading matrix. In the extreme case, all variables are just linear combinations of one stochastic trend, and the 'long-run' equilibrium is realized almost immediately after a shock. The performance of the tests might depend on how strongly the variables are tied to the long-run relation. In a set of experiments not reported here, where cointegration takes that form, both the LLC and the IPS test were very badly oversized. Since in that setup a common stochastic trend is a time specific effect common to all series, step one of LLC just eliminates it and transforms all series into stationary processes. Lower values of the loading matrix $\mathbf{A}$ may seem more realistic and loosen the tightness of the long-run relation. Here the DGP takes the following form:

$$\mathbf{A} = 0.1 \begin{pmatrix} -\mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} \mathbf{I} & -\mathbf{I} \\ 0 & \mathbf{C} \end{pmatrix}$$

the cointegrating matrix $\mathbf{C}$ is:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 \\ & & & & \dots & \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

The number of zero rows (b) determines the number of common trends driving the system. There are $N-b$ cointegrating relationships. The following number of cointegrating vectors were considered: $N-1$, $N/2$ and $N/4$, in case of a fraction the integer part of it is chosen.

Table 5.6: Size properties with cointegration, $b = N - 1$

$b = N - 1$	LLC			IPS		
	nominal size 10%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.167	.157	.158	.161	.150	.164
$T = 50$	.192	.167	.173	.261	.255	.274
$T = 100$	.249	.227	.203	.520	.479	.449
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.094	.078	.088	.096	.086	.096
$T = 50$	.105	.081	.088	.153	.147	.171
$T = 100$	.139	.132	.117	.383	.378	.356
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.024	.019	.024	.030	.024	.026
$T = 50$	.024	.017	.017	.044	.041	.055
$T = 100$	.031	.029	.034	.159	.182	.198

Note: Based on 4,000 replications.

Table 5.6 reports the results for $CIV = N - 1$. The tests are oversized and the problem increases in T. Together with Tables C.2 and C.3 in the Appendix, where results for other values of CIV are presented, it becomes clear that the tests perform worse the more CIVs are present.

An analytical treatment of the asymptotic behaviour of the LLC test statistics for the cases considered in the simulation exercise would give insights into the origins of the size distortions. The interested reader is referred to Lyhagen (2000), who provides an analytical argument for the special case in which there are $N-1$ cointegrating relations and an instantaneous adjustment to the equilibrium takes place ($\alpha = 1$). He derives the limiting distributions for the t-statistic. The variety of parameters that can determine the cointegration among the sections (number of CIVs, α) makes a general analytical treatment of this bias rather complicated. Furthermore, the additional insight of an analytical treatment is limited as a potential correction of the size distortion would have to account for all possible cases.

5.3.3 Cross-Sectional Correlation and Cointegration

The two previous sections indicated that both kinds of dependencies have oversizing effects and therefore yield to an over-rejection of the null-hypothesis. Neither econometric nor economic theory gives any reason to believe that the two dependencies are mutually exclusive. In this sections the two are brought together. The cointegration is chosen to be as in the above section, and, in addition, the errors are correlated in the way specified in Section 5.3.1.

The results reported in Table 5.7 are surprising. The distribution of the test statistic is shifted to the right. The bias increases with T and yields a considerable distortion in the opposite direction. This of course causes the power of the test to come close to unity.

In theory, corrections to the test statistics are possible. The variety of cases (N, T, number of CIVs, α), however, limits the practicability of such an approach. Hence, in practice, a careful assessment of the dependencies present in the data is necessary before applying any unit root test.

5.4 Should Panel Unit Root Tests be Applied to Exchange Rates?

With the growth of the panel unit root methodology, the debate over the validity of the purchasing power parity (PPP) has experienced a revival. While previous research could hardly find any empirical evidence for PPP, one could

Table 5.7: Size properties with cointegration and correlation

$CIV = N/2$	LLC			IPS		
$\omega = 0.7$	nominal size 10%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.092	.055	.029	.065	.036	.015
$T=50$	.094	.058	.029	.054	.035	.007
$T=100$	.111	.064	.025	.061	.024	.004
	nominal size 5%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.049	.030	.012	.038	.019	.007
$T=50$	.055	.033	.014	.030	.010	.004
$T=100$	.067	.039	.012	.035	.010	.002
	nominal size 1%					
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
$T=25$	.011	.007	.002	.011	.003	.001
$T=50$	.018	.001	.003	.007	.001	.001
$T=100$	.032	.013	.002	.011	.004	.000

Note: Based on 4,000 replications.

expect more insight from the application of panel methods.[6] In a non technical way, PPP means that once different currencies are controlled for, the same basket of goods should cost the same amount of money no matter in which country it is purchased. The existence of permanent deviations from such an equilibrium seems implausible as it would allow arbitrage gains, which in turn would push the real exchange rate back to the equilibrium. Although nobody believes in arbitrage possibilities with fast food, a popular application of PPP is the Economist's Big Mac index. Assuming PPP holds, actual exchange rates are expressed as the deviation from the McParity, hinting on the current under- or overvaluation of currencies.[7] Although many arguments have been put forward in the theoretical literature why PPP might fail, PPP is still a very popular concept and something many economists like to believe in. However, one cannot reject the impression that much of the debate centres on the applied methods.

6 For a survey of empirical results before the panel era, see e.g. Froot and Rogoff (1995).
7 The fall of the Euro after its introduction was predictable if one had believed in Burgernomics. For more on the issue, see Economist (2001, April 21st).

5.4.1 PPP - Revisited

If PPP holds, in the long run the real exchange rate between two countries is stable and deviations from equilibrium are not permanent. Let E_{it} denote the nominal exchange rage between country i and a *base* country at time t. Then, multiplying a basket of goods (normalized to one) with the ratio of the prices in country i, P_{it}, and in the country of the base currency, P_t^{base}, defines the real exchange rate Q_{it}:

$$Q_{it} = \frac{1}{E_{it}} \frac{P_{it}}{P_t^{base}}$$

or, taking logs:

$$q_{it} = p_{it} - p_t^{base} - e_{it}. \tag{5.8}$$

Since prices and exchange rates are recognized to be nonstationary time series, a natural way of looking at the problem is to ask if there is a linear combination of the series which renders a stationary real exchange rate, *i.e.* if the prices and the exchange rate are cointegrated.

A distinction is made between the strong and weak form of PPP. The weak form allows for coefficients different from (1, -1) on the price indices. The weak form of PPP has its economic justification in the presence of measurement errors, which would persist in the long run, or varying effects of productivity shocks which may cause the cointegrating coefficients to differ from unity. The weak form of PPP has been tested in an error correction approach, *e.g.* by Cheung and Lai (1993) or Corbae and Ouliaris (1991). Edison et al. (1997) and Kouretas (1997) apply a Johansen (1995) procedure, the latter to investigate PPP of the Canadian dollar and five other currencies.

The strong form of PPP, restricts the coefficients to (1, -1) *a priori* and tests the resulting real exchange rate for a unit root. Only this test is of interest in the panel unit root framework. The PPP hypothesis translates into the stationarity of the real exchange rate q_{it}: Only if this series is mean reverting and does not accumulate shocks permanently, can PPP hold. Interestingly, the majority of the studies apply tests that have a unit root as a null hypothesis and literally *accept* stationarity if nonstationarity is rejected, which clearly is a loose interpretation of the unit root rejection. Kouretas (1997) and Kuo and Mikkola (1999) are two studies which test both stationarity and nonstationarity in a panel framework. In a univariate framework, Engel (2000) points out that even if one rejects the unit root and fails to reject stationarity there is a possibility of a unit root in the series. This might be caused by a size distortion in the unit roots tests and the low power of the stationarity tests.

Several issues make PPP an interesting application from the perspective of panel unit root tests. The increased power when taking into account a set of time series allows for a more precise statement on the stationarity of the series. While single country analyses often reject PPP because a unit root is found in the real exchange rate, this might be due to the low power of single equation unit root tests with an autoregressive coefficient close to unity. Therefore, the panel approach might give more insights. However, there are drawbacks on the use of panel methods. Interestingly, some authors find differing results according to the base currency chosen. Papell (1997) rejects the unit root when the Deutschmark is chosen as a base currency, but has mixed results when the panel is US$ based. Note that the series to be tested for a unit root formed following equation (5.8) exhibit cross-sectional correlation by construction, as they are expressed with respect to one base currency. Hence, shocks that affect this exchange rate are directly reflected in the entire panel. This means that the degree of cross-sectional correlation depends on the base currency chosen. However, Tables 5.3 and 5.4 in Section 5.3.1 show that the actual value of the cross-sectional correlation does not influence the performance of the test. It is more plausible that the choice of the base currency affects the degree to which the data is contaminated with cointegration.

There is, of course, a debate on what long-run means in this context. While some authors argue that PPP should hold regardless to the exchange rate regime, and consequently apply the tests to long series from, say 1949-1996 (Kuo and Mikkola, 1999), or even over 100 years (Engel, 2000), most of the studies rely on the time period of the current float, *i.e.* from 1973 onwards.

All studies mentioned above, including Pedroni (1999), do not consider the possibility of cross-sectional cointegration. Banerjee et al. (2001) confirm the result of the previous cointegration analysis that if cross-sectional cointegration is not taken into account when the real exchange rate is computed, severe distortions may arise. Although one should be aware of the possibility of cross-sectional cointegrating relations and the serious distortions this causes, one has to recognize that large dimensional systems cannot be estimated without an *a priori* restriction. To illustrate this argument, a full Johansen estimation of the weak form of PPP would yield a system of N countries, each with 3 variables, so that an unrestricted estimation of the cointegration matrix Π would not be feasible with some 100 observations.

As mentioned earlier, the study by O'Connell (1998) examines PPP in the presence of short-run dependencies in the form of cross-sectional correlations. Moreover, the size and power of the LLC test are explicitly analyzed. O'Connell comes to the conclusion that the performance of the LLC test in the presence of cross-sectional correlation is very poor and suggests a new

GLS type estimator. The impact of the O'Connell critique was considerable and has to some extent discredited the LLC test. There are some things worthwhile noticing.

Apparently O'Connell does not use the adjusted t-value when he evaluates his simulations results. Not adjusting the t-values means that the finite sample adjustments are not made. Also, common time effects are not eliminated in his simulation exercise. This becomes clear as the distortion in size he reports can only be reproduced if one does not perform this elimination. The poor power properties that are attested to the LLC test are not related to the cross-sectional correlation (see Table 5.3). It should be pointed out that with a specification of $\rho = 0.96$ even univariate unit root tests have poor power results (Schwert, 1989). Thus the poor power properties are not panel specific. The proposed GLS estimate may seem more appealing than the removal of common time effects. However, this procedure involves the estimation of a covariance matrix and relies on the consistency and accuracy of this point estimate.

5.4.2 Shortrun Dependence

The main finding of the simulation exercise above is that it is essential to know more about the covariance structure of the data before applying unit roots tests. This poses some methodological problems because estimators have to deal with possible heteroscedasticity and serial correlation in the data. Robust estimators are needed. In addition, once a point estimate of a covariance matrix is obtained, it is necessary to conduct some inference on the parameters in order to asses the significance of the correlations. Parametric (Den Haan and Levin, 1996) and nonparametric (Newey and West, 1987) methods for robust estimations of covariance matrices are discussed in Appendix C.2. In addition, a bootstrap algorithm (Politis and Romano, 1994) is suggested to test for significance of the estimated correlations. The data used are a panel of real exchange rates for 18 OECD countries, using the US$ as the base currency.[8] To be consistent with the covariance estimators that operate under stationarity, the first difference of the real exchange rates form the basis for the following analysis. This is consistent while working under the null hypothesis of a unit root. The parametric (Table 5.9) and the nonparametric (Table C.6 in the Appendix) estimation yield similar results for the covariance matrix and show clear signs of significant positive correlation.[9]

8 The data used is from the IMF data sets, namely the International Financial Statistics and covers quarterly nominal spot exchange rates and CPI, for the period 1973:1 to 1997:3 for 18 OECD countries. A plot of the data can be found in the Appendix.

9 The estimates presented are not sensitive to the choice of parameters (information criterion, lag lengths, truncation $\bar{K}$). Although the two estimates are not identical, their

One can argue that the parametric estimate is superior because it explicitly considers prewhitening which is the main drawback of the nonparametric estimator used. On the other hand the differenced exchange rates do not, in general, have very high order autoregressive components,[10] so the impact of serial correlation on the nonparametric estimator might be limited.

More than the nonparametric estimate, the parametric estimates detect negative correlation of the Canadian Dollar with most other currencies in the point estimates. However, the standard errors indicate that it is not significant. Recalling that all variables are constructed the following way: $\mathbf{x}_{1t} = \frac{GBP}{US\$}, \mathbf{x}_{2t} = \frac{ATS}{US\$}$ and that from the 18 countries chosen most are European, it is not surprising that Canada seems to react in a different way to shocks – if affected at all. The same is true for Korea. The Japanese Yen, on the other hand, does exhibit similar reactions to the European currencies. In the parametric case, the standard errors are in a plausible range of 0.02 to 0.2, whereas in the nonparametric case, the standard errors become very small, especially if the estimated correlation is close to unity. Overall, the parametric estimation seems more plausible.

Having in mind the results from the simulation and the asymptotic considerations of Section 5.3, it is desirable to have a homogeneous dataset in terms of error correlation. Therefore, the two countries with a different error correlation (Canada and Korea) were dropped from the sample yielding a panel with almost equally correlated errors. The tests on different structures of the the covariance matrix of the remaining 16 countries reported in Table 5.8 indicate that a common correlation coefficient in the order of .6 to .8 cannot be rejected, with 0.7 yielding an exceptionally low test statistic.

Table 5.8: Testing different covariance structures for 16 currencies

	$p = 0.25$	
ω	Q	p-val
0.5	295.82	1.00
0.6	85.75	0.01
0.7	2.06	0.00
0.8	44.76	0.00
0.9	213.84	1.00

Note: Q is χ^2_{120} distributed.

results are very similar, and the deviations from each other are in a plausible range (see, e.g.[Section 6]Den Haan and Levin (1997)).

10 The average lag length is 2.7, with a range from zero to 7 in one case.

Table 5.9: Estimated sectional correlation in the differenced exchange rates

	UK	AT	BE	DK	FR	DE	NL	CA	JP	FN	GR	ES	AU	IT	CH	KO	NW	SW
UK	1.00																	
AT	0.55 0.09	1.00																
BE	0.59 0.08	0.96 0.02	1.00															
DK	0.62 0.07	0.92 0.03	0.94 0.02	1.00														
FR	0.62 0.07	0.89 0.04	0.89 0.04	0.88 0.04	1.00													
DE	0.59 0.08	0.99 0.01	0.97 0.02	0.94 0.02	0.90 0.03	1.00												
NL	0.63 0.07	0.97 0.02	0.96 0.02	0.93 0.03	0.88 0.05	0.98 0.01	1.00											
CA	0.06 0.10	−0.03 0.11	−0.04 0.11	0.02 0.11	−0.01 0.10	−0.03 0.11	−0.04 0.11	1.00										
JP	0.50 0.08	0.62 0.07	0.63 0.07	0.64 0.07	0.61 0.08	0.63 0.07	0.64 0.07	−0.01 0.11	1.00									
FN	0.64 0.08	0.64 0.12	0.65 0.11	0.63 0.12	0.60 0.13	0.65 0.10	0.66 0.10	0.11 0.11	0.44 0.09	1.00								
GR	0.65 0.07	0.75 0.12	0.77 0.13	0.79 0.11	0.79 0.11	0.79 0.13	0.79 0.12	−0.01 0.14	0.55 0.07	0.58 0.08	1.00							
ES	0.63 0.08	0.74 0.07	0.73 0.07	0.73 0.07	0.77 0.07	0.74 0.09	0.75 0.07	−0.04 0.12	0.51 0.11	0.58 0.11	0.65 0.07	1.00						
AU	0.21 0.12	0.22 0.10	0.24 0.10	0.28 0.10	0.28 0.10	0.22 0.10	0.25 0.10	0.27 0.10	0.24 0.12	0.21 0.11	0.33 0.10	0.22 0.09	1.00					
IT	0.62 0.08	0.69 0.09	0.72 0.08	0.72 0.09	0.81 0.07	0.71 0.09	0.71 0.09	−0.04 0.13	0.50 0.09	0.59 0.12	0.69 0.09	0.74 0.08	0.23 0.08	1.00				
CH	0.53 0.09	0.84 0.04	0.82 0.05	0.82 0.05	0.87 0.05	0.85 0.04	0.82 0.06	0.05 0.12	0.64 0.06	0.57 0.13	0.70 0.09	0.66 0.10	0.23 0.10	0.70 0.08	1.00			
KO	−0.07 0.09	0.05 0.12	0.03 0.12	0.03 0.12	0.09 0.11	0.05 0.11	0.03 0.11	0.22 0.09	0.13 0.09	0.03 0.11	0.01 0.09	0.10 0.09	0.29 0.14	0.08 0.09	0.09 0.15	1.00		
NW	0.60 0.10	0.85 0.03	0.82 0.04	0.79 0.06	0.85 0.05	0.85 0.03	0.81 0.04	−0.01 0.12	0.52 0.08	0.64 0.13	0.73 0.10	0.74 0.08	0.30 0.09	0.68 0.09	0.76 0.05	0.09 0.09	1.00	
SW	0.61 0.08	0.71 0.07	0.69 0.08	0.68 0.09	0.68 0.08	0.70 0.08	0.69 0.07	−0.02 0.11	0.42 0.11	0.70 0.09	0.60 0.12	0.76 0.09	0.21 0.09	0.68 0.09	0.62 0.09	0.05 0.12	0.77 0.06	1.00
	UK	AT	BE	DK	FR	DE	NL	CA	JP	FN	GR	ES	AU	IT	CH	KO	NW	SW

Note: The values reported are correlation coefficients obtained after the correlation transformation of the estimator in equation (C.3). The standard errors reported are based on 1000 replications of the bootstrap procedure described in Section C.3 and computed as in (C.5).

5.4.3 Long-run Dependence

A full assessment of the long-run dependency in the real exchange rate data is not possible using a maximum likelihood approach due to the few numbers of observations in relation to the entire system. A $VAR(p)$ specification of the process that satisfies minimal residual properties would require a lag order higher than $p = 2$, which is the highest feasible in the system of 16 exchange rates. Estimation might be achievable by imposing further *a priori* restrictions on the parameter matrices, but theory does not give any further guidance. However, the interesting question whether there is cointegration or not can positively be answered in subsystems of the 16 exchange rates. For the sake of presentation, here the result of a sub-sample of 9 exchange rates is presented.[11] A cointegration analysis following Johansen (1995) suggests that the data are cointegrated. The trace test detects at least three cointegrating relations in this sub-sample of the data (see Table 5.10). This exercise could be repeated with varying sub-samples yielding similar results.

Table 5.10: Trace test for cointegration in a sub-sample of 9 exchange rates

H_0: rank=p	$-T\sum \log(1-\hat{\lambda}_i)$	95%
$p = 0$	234.5**	192.9
$p \leq 1$	183.3**	156.0
$p \leq 2$	136.7**	124.2
$p \leq 3$	97.49*	94.2
$p \leq 4$	65.67	68.5

Note: ** indicates that the hypothesis is rejected at least at the 95%-level.

5.4.4 Results

The individual lags that were included in the different sections were determined after the removal of common time effects. This lag structure differs from the optimal lag structure if each of the series would be tested individually before demeaning. However, because the absence of serial correlation

11 A $VAR(3)$ was fitted allowing for seasonal dummies and a constant. All Box Pierce statistics testing for the absence of serial autocorrelation up to 11 lags cannot be rejected, the same is true for ARCH(4) effects. Absence of vector autocorrelation is rejected at the 5 percent level, vector normality is not rejected.

is essential for the LLC and IPS test this should be carried out after the demeaning. A series of tests was used to analyze the residuals of each series for their white noise properties. Table 5.11 reports the results and the main white noise indicators. Further, the t-values of the included lags were considered which had to be significant at least for the highest lag considered. Normality is not rejected for all residual series.

Table 5.11: ADF - lags selection

	US$ based						
	lags	AR(4)	BP		lags	AR(4)	BP
DE	3	.59	.99	FN	4	.25	.58
UK	1	.90	.34	GR	5	.87	.71
AT	1	.57	.87	ES	1	.78	.60
BE	3	.77	.24	AU	3	.96	.84
DK	2	.44	.45	IT	2	.80	.85
FR	1	.74	.83	CH	1	.54	.85
NL	3	.47	.90	NW	3	.87	.12
JP	5	.41	.60	SW	2	.84	.85

Note: In all cases the null hypothesis is absence of the respective disturbance. The reported values are the p values at which this hypothesis can be rejected. AR(4) stands for a test on autocorrelation to the 4th order, BP is the Breush Pagan test for heteroscedasticity.

Both dependencies are present in the data. The simulation exercise has shown that in this case it is not possible to make predictions about the direction of a potential size bias. New critical values can be computed simulating panels of exchange rates, and thereby following as close as possible the presumed DGP. Therefore, a panel of 16 variables with 15 (and 8) cointegrating relations and an error correlation structure using the point estimate of Table 5.9 was simulated and the test statistic was computed. Under the alternative, variables with an autoregressive coefficient of $\rho = \{.9, .95\}$ and the same error structure were simulated. Table 5.12 reports the results of the LLC and the IPS test, the percentiles of the normal and the simulated distribution. In addition, Figure 5.1 shows kernel densities of the estimated t-values, the standard normal distribution, the actual test value and the $\rho = 0.9$ alternative.

Table 5.12: Test results

test	coefficient	p–values N(0,1)	simulated $b=15$	simulated $b=8$	adj. power (5 %) $\rho=.9$	adj. power (5 %) $\rho=.95$
LLC t^*	-1.911	0.028	0.014	0.020	0.94	0.38
IPS $\Psi_{\bar{t}}$	-2.856	0.002	0.001	0.002	0.99	0.73

Note: Simulated values based on 4,000 replications.

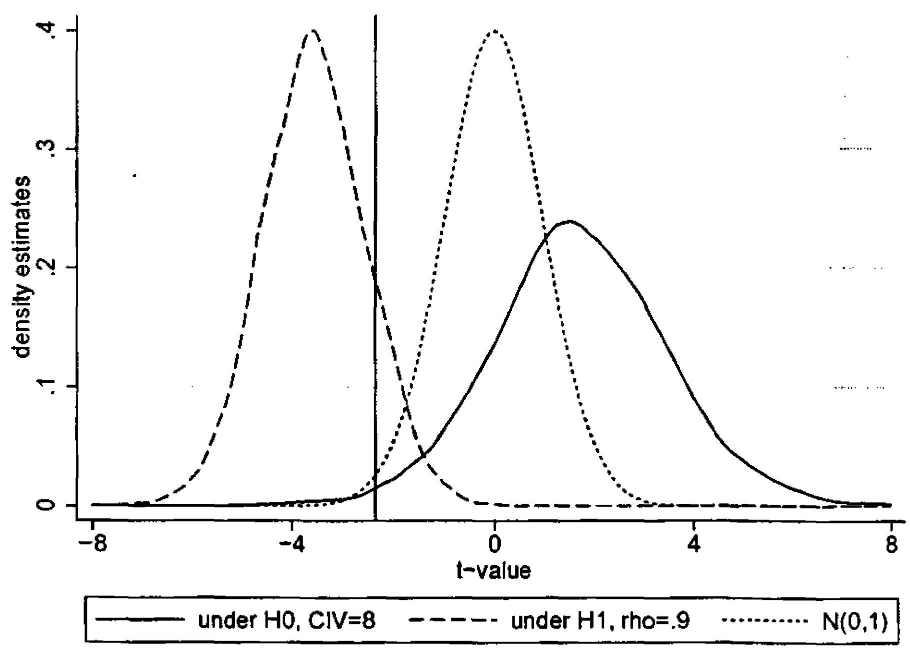


Figure 5.1: LLC test results.

One has to keep in mind, however, that this test result is sensitive to the assumed structure of the data, in particular the presence of both long- and short-run dependence. While the short-run correlation with all errors sharing the same correlation coefficient does not appear to influence the test result, the presence of cointegration is much more worrisome. If one is willing to assume that the values obtained via the simulation reflect the true properties of the DGP, the null of nonstationarity of the real exchange rates can be rejected at a very low p-value, hence providing some argument for the validity of PPP.

5.5 Conclusion

The simulation exercise has shown that two of the most popular panel unit root tests are sensitive to dependencies among sections of the panel. Analyzed separately, both short-run dependence in the form of correlated errors and long-run dependence in the form of cointegration lead to a significant oversizing of the test. However, if put together, the effect goes into the opposite direction. The determination of the actual presence of dependencies is therefore necessary in order to interpret the test results. The estimation of and the inference on contemporaneous correlation is crucial, although not easy to perform.

The application to a set of real exchange rates has shown that both dependencies are present in the data. Hence, the test results are likely to be biased. In order account for these dependencies, simulated critical values were used which origin from a data generating process that resembles the actual data. The null hypothesis of a unit root can be rejected providing some empirical evidence for the validity of the purchasing power parity theory. However, in the light of the simulation results obtained earlier, the reliability of the test results are questionable. This exemplifies the the problems with the use of panel unit root tests on sectional dependent data in general.

Appendices

A Appendix to Part I

A.1 Experimental Design

The experiment was conducted using a computerized setup[12] in three sessions at the European University Institute near Florence, Italy. Participants were 110 Masters and PhD students from the faculties of Law (30%), History (15%), Social and Political Sciences (23%), and Economics (33%). Subjects originated from 15 different European countries. They were between 23 and 36 years old (average: 27.7), and 64% were male. Because it was the first time that experiments were conducted at this place, the subject pool was not experienced in playing games. For each session a multiple of five subjects was recruited (session 1: 40, session 2: 30, session 3: 40). The profit earned by participants ranged from Euro 24 to Euro 47.90, with an average of Euro 36.34 (s.d. 4.89), including a 5 Euro show-up fee paid to each candidate. Each session (including a 15 min. questionnaire at the end) lasted for about two hours. Participants were recruited via email and were invited to sign up on a website. Each session took place in three computer labs with 10 to 25 computers each, located in different buildings of the university campus. Upon arrival to an assigned computer lab, subjects randomly drew a seat number and an account number. This account number was later used to identify subjects for payment, which was organized anonymously. Further to that, the computer labs were prepared using separators to individualize the environment. In each room, a professor of the university monitored the experiment in a discrete way.

Note that at no point in time were subjects deceived. Subjects could choose how often (max three times) they wanted to read through the instructions on the screen. They also had a hard copy of the instructions next to their machines. The instructions were followed by a short quiz of three questions covering the crucial aspects of the game. Almost all subjects appeared to have understood the game very well before playing (see results of the quiz in Section A.1). No major clarification questions were asked. After reading through the instructions, subjects were asked to enter information about their age, gender, nationality, and number of siblings.[13] To increase anonymity, the age displayed to fellow players was modified by adding a random number. This was also mentioned in the instructions further to a general anonymity and privacy statement.

12 The Z-Tree software described in Fischbacher (1999)

13 During the recruitment process it was made sure that subjects were recruited only from countries which have a substantial number of students at the university. This restriction was introduced to avoid identification of the subjects during the game.

Each session consisted of six treatments in which subjects were randomly matched in groups of five players

In the first four of these treatments subjects played the following repeated version of the trust game. At the beginning of the treatment, each player could see some information about the four other players in his group. the information included the players' nationality, age, gender, and the number of siblings. The subjects then decided to *whom* and *how much* of their initial endowment of 100 they were willing to transfer. No entry in any of the boxes corresponded to making no choice, which was also an option. In the next step subjects saw *who* among the other players had chosen them and *how much* they had received from these partners. In addition, this amount was shown multiplied by three. For each player from whom a transfer was received, they could choose how much to return back. Then, subjects were presented a summary of all transfers and returns they had been involved with. These steps were repeated six times. Then, groups were reshuffled and a new treatment was played. Due to the limited amount of subjects in each session and the large size of each group, the re-matching had to be done on a random basis, hence it is not ruled out that subjects could meet again in subsequent groups.

The fifth treatment differed from the previous four because it did not allow free choice of partners. Subjects, were again matched in groups of five players, but instead of being able to choose a partner, they were randomly assigned to one of the fellow players. In every period of this treatment players faced a new non-modifiable random choice of partner.

The sixth and final treatment was instead identical to the first four and thus allowed free choice of partners.

Instructions

Screen 1

- You will randomly be matched with **4 other players** to play a **game**.
- Each game consists of **three stages** which will be described on the following screens.
- The game will be repeated **for 6 periods** with the **same** players.
- After the 6 periods, you will randomly be **re-matched** with four new players.
- This re-matching will be repeated **six times** (time permitting).

Screen 2

Stage 1 of 3

Your endowment in each period is **100 points**, equivalent to **0.35 Euros** You can choose **if** you want to transfer any points to your fellow players or not. If so, you decide **to whom** and **how much**. You can choose **only one person** and you can transfer any amount between **0 and 100**. If you decided not to transfer points at all, just click the button.
Every transfer made in stage 1 will be **multiplied by the factor 3** as it arrives on the other player's account.

Screen 3

In stage 1 the other 4 players have simultaneously made a similar decision to yours. Due to the simultaneity their choice does not depend on your decision.

Stage 2 of 3

You will see **who** of the other players have chosen you and **how much** has been transferred to you. It might be that you were chosen by none, 1, 2, 3 or even all 4 players.
If you got a transfer from a player, you can decide **if** and **how much** you want to transfer back to **this player**. You can transfer back anything **between zero and three times** the initial transfer to you. If you were chosen by more than one player, you can choose different amounts for each of them.

Screen 4

Stage 3 of 3

In this stage you see the results of the period, how much you transferred and how much the player you have chosen initially **transferred back** to you. You will also see the profit in Euro you made in this period.

Screen 5

Remember...

- After you finished playing the three stages, you will play this game six times with the same players.
- After the 6 periods, you will randomly be **re-matched** with four new players.
- This re-matching will be repeated **six times** (time permitting).

Do you want to read the instructions again or continue directly with a short quiz?

Screen Before the Predetermined Treatment

The game you will play now is **slightly different** from the one you have played before.
Contrary to the previous game, in Stage 1 you will **not have the possibility to choose a player**. Instead, a **random choice** will be made for you. You can only **decide how much** you want to transfer to the player already determined.
Notice that this also affects stage 2, as it is now random by how many players you were chosen.

Privacy Statement

The privacy and Anonymity statement reads as follows.
All information we collect undergoes a strict anonymization process, not only ensuring anonymity among players but also ensuring that you stay anonymous to us. No private information will be collected. During the experiment you will see some information about your fellow players. We have ensured that you cannot identify them personally, and vice versa, they cannot identify you. Remember that this experiment runs over different rooms, thus involving much more individuals than those seated in your room. At the end of the session, you will be asked to type in the account number you obtained before. Please keep this number, because after notification you can pick up an envelope with your payment at the porters' lodge.

Screenshots

See Figures A.1 to A.5 for some screenshots of the game.

Figure A.1: Screenshot of the information collected at the beginning

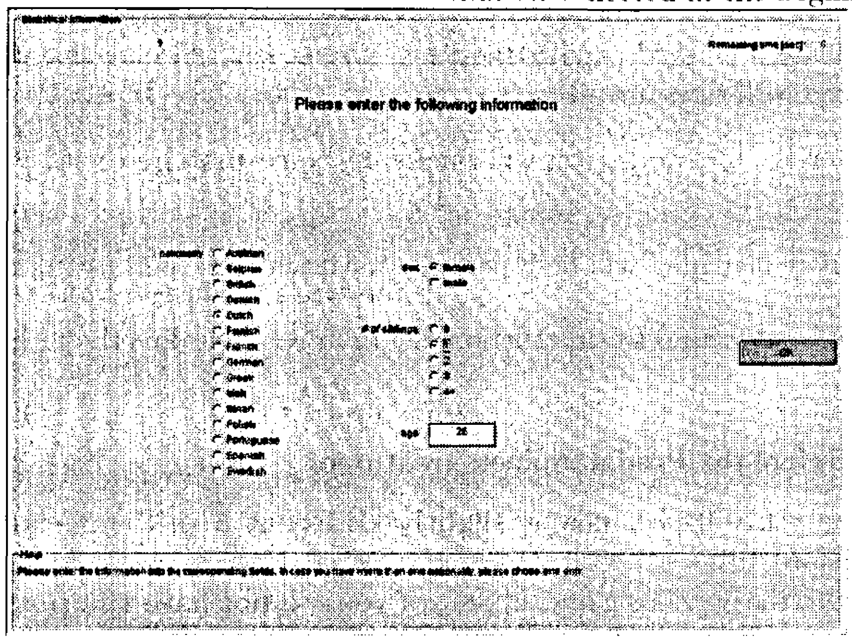


Figure A.2: Screenshot of the first stage

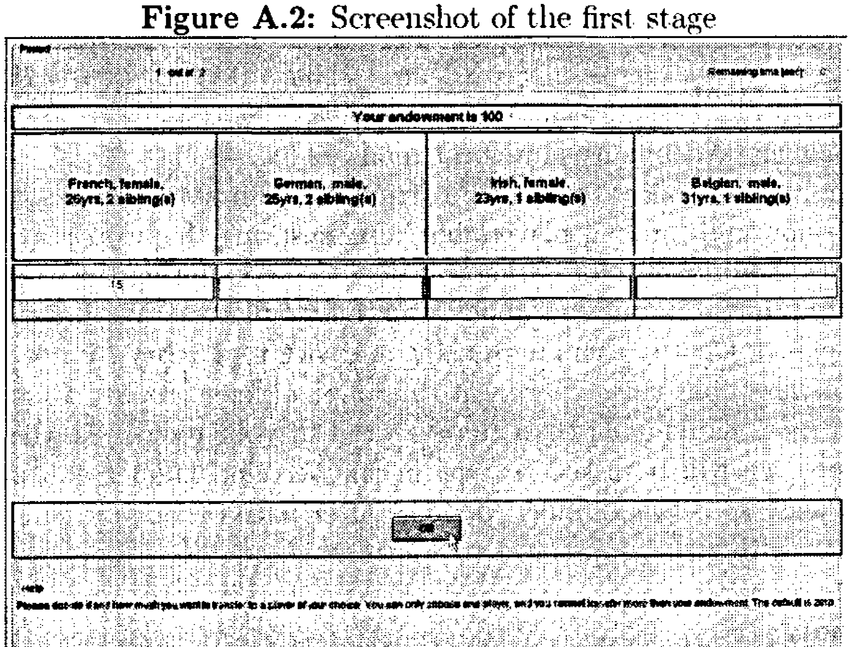


Figure A.3: Screenshot of the second stage

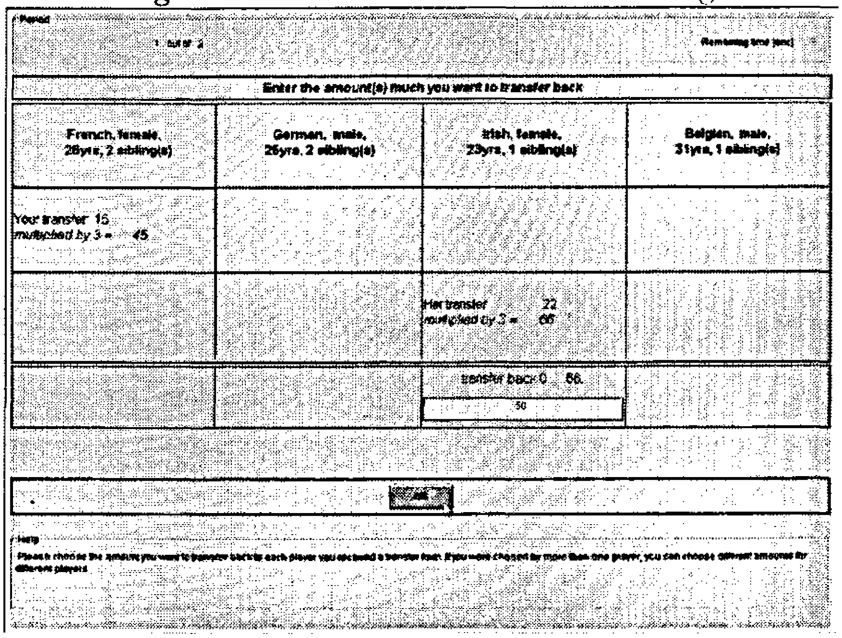


Figure A.4: Screenshot of the third stage

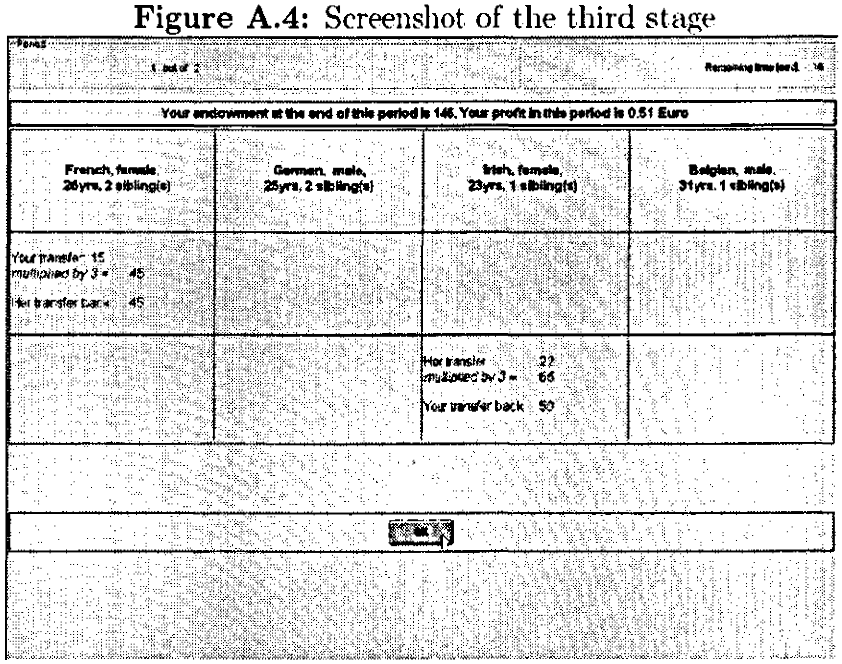
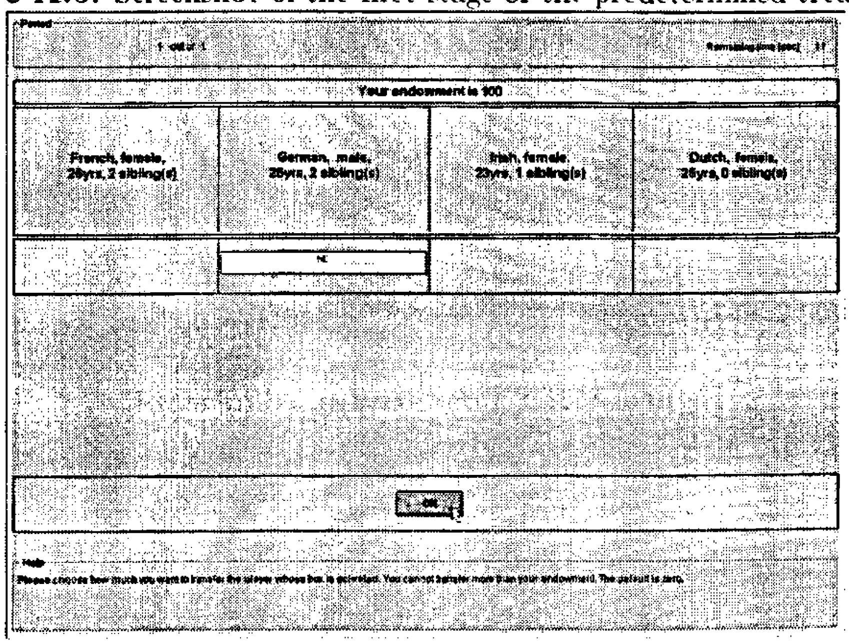


Figure A.5: Screenshot of the first stage of the predetermined treatment



Quiz

Note: Subjects always saw the actual values of the expressions involving X, Y, Z.

Question 1: [Subjects had to choose an amount X between 1 and 100.] "Imagine you transferred X points to player two in stage 1. Assume further that she made no transfer to you in stage 1. How many points can you transfer back to her in stage 2 at most?"

A: $3X$ B: X C: 0

Question 2: [Subjects drew random number Y between 0 and 100 by clicking on a button.] "Your drew the number Y. Assume you transferred this amount to one player in stage one. How much can the other player transfer back to you at most?"

A: 0 B: $3Y$ C: Y

Question 3: "Please press the button below to determine randomly how much you will be paid back. Remember that this number can be between 0 and $3Y$." [next screen] "Summary Question: Initially, from your 100 points you transferred Y to the player. Let us assume the player transferred you back Z in the next stage. You had no interaction with other players. Based on this, what is the balance on your account?"

A: 0 B: $3Y$ C: $100 - Y + Z$

Table A.1 summarizes the results. Subjects got a feedback screen after each answer indicating if they were correct or mistaken and stating the correct answer. While in the first question many subjects made mistakes, in questions 2 and 3 almost all subject answered correctly.

Table A.1: Results of the quiz, in per cent

Answer	Question 1	Question 2	Question 3
A	19	1	1
B	21	95*	5
C	60*	4	94*

Note: * denotes the correct answer.

A.2 Additional Figures

Figure A.6: Histogram of transfer, periodwise

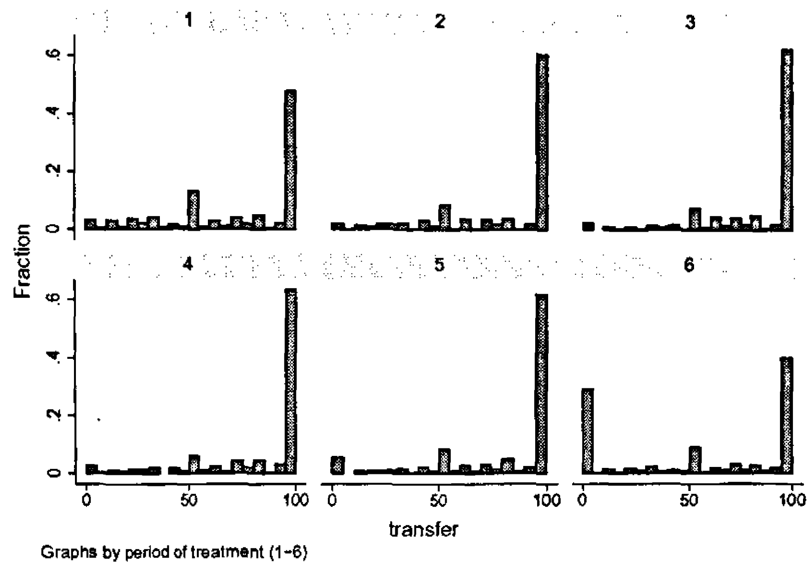


Figure A.7: Histogram of returned ratio, periodwise

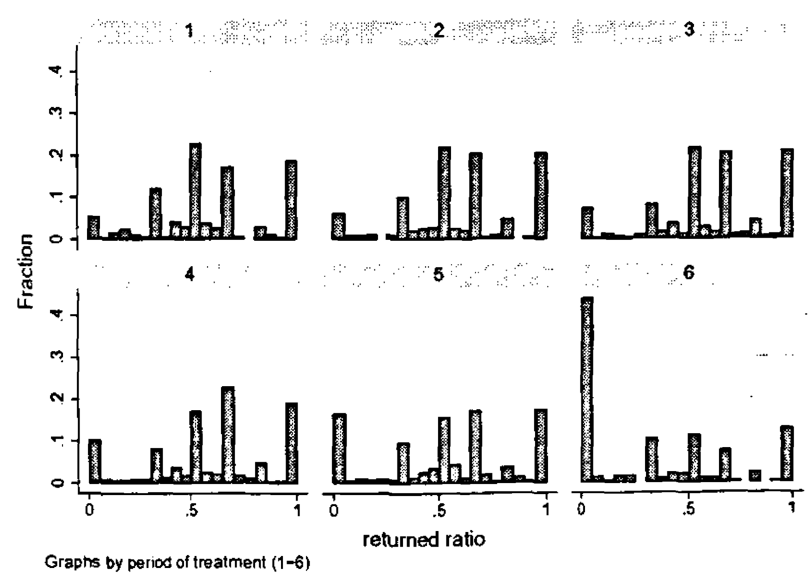


Figure A.8: Histogram of transfer, predetermined treatment, periodwise

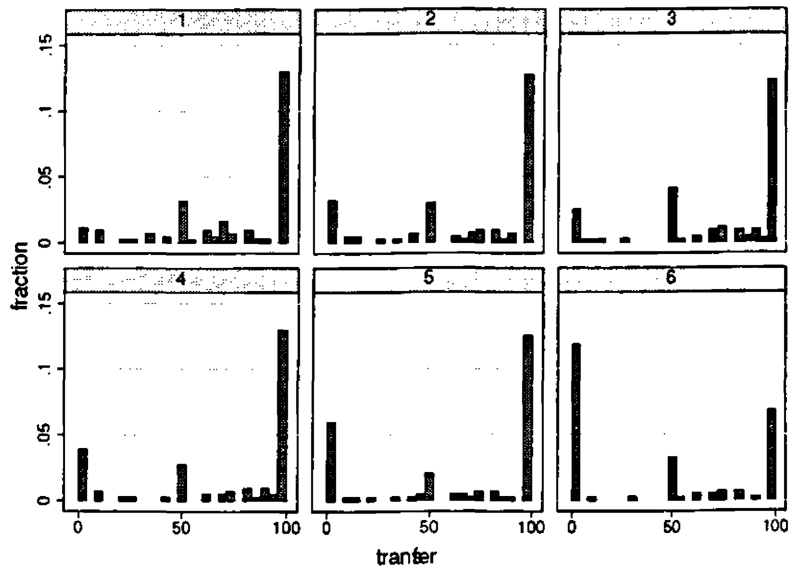
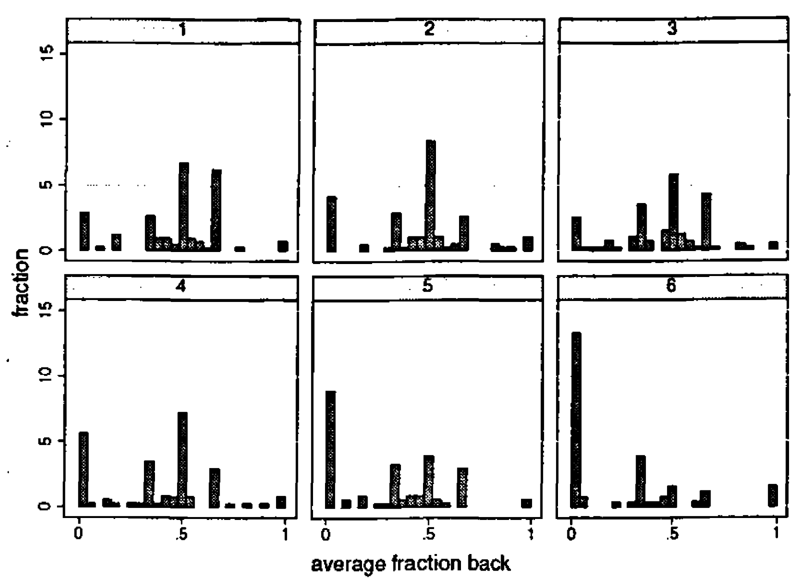


Figure A.9: Histogram of returned ratio, predetermined treatment, periodwise



B Appendix to Chapter 4

B.1 Key Elements of PROGRESA

PROGRESA is the acronym of *Programa de Educación, Salud y Alimentación* and is a programme that aims at developing the human capital of people living in poor rural households in Mexico.[14] Launched by the federal government of Mexico in 1997, the International Food and Policy Research Institute (IFPRI) joined the effort a year later together with other research and development institutions. The programme is still running band is now known as OPORTUNIDADES.

One of the main objectives of the programme is to improve the school attendance of children. Eligible households with school aged children receive grants conditional on school attendance. The size of the grant increases with the grade (starting from the third year of primary school) and, for secondary education, is slightly higher for girls than for boys (see Table B.1).

Initially, 506 localities were chosen to participate in the programme. For logistic reasons and evaluation purposes, the sample was divided into to a treatment (320) and a control (186) group, where the programme started two years later. The selection into the treatment and control group can be considered as having been random - at least with respect to the variables that interest in this analysis. Within each village the survey covers all households (roughly 24,000 observations) and collects extensive information on consumption, income, nutrition and other issues. For each *household member*, including each child, there is information about age, gender, education, labor supply, income (various forms), school enrolment, nutrition, and health status. Detailed information about the localities is also available. However, the questionnaire used varies substantially between waves.

Based on the information collected in the first round of interviews an eligibility criterion was established and the sample was classified into eligible and non eligible households. Later the eligibility was extended (known as the *densificación*) such that now finally some 80 percent of the households were eligible to participate in the programme.

14 For a detailed description of the programme and an extensive documentation, see www.ifpri.org/themes/progresa.htm.

Table B.1: PROGRESA transfer scheme

Priamry school		Secondary school		
			transfer	
grade	transfer	grade	boys	girls
3^{rd}	130	1^{st}	380	400
4^{th}	150	2^{nd}	400	440
5^{th}	190	3^{rd}	420	480
6^{th}	260			

Note: Values indicate bimonthly transfers in Mexican pesos payable to each eligible child on attendance of the specific grade (values valid for the first semester 1998). 10 pesos roughly corresponded to 1.10 US Dollar in that period. A maximum of 1,170 pesos was payable to each household.

B.2 Data

The data of the first five waves is publicly available from the website of IFPRI.[15]

Table B.2: Sample means by reported status

Variable	not school	work and school	school only
age	14.45	11.85	12.28
female	.52	.41	.47
years of schooling	5.74	4.61	5.06
rank	1.73	2.13	2.19
children below 6 in hh	.87	.94	.88
median child wage in state	90.0	36.23	41.68
number of total hh members	7.31	6.96	7.20
weekly per capita household income	53.93	53.21	42.56
schooling of most educated parent	2.65	3.18	3.39
age of the oldest parent	46.27	43.93	45.03
share of control	.40	.34	.38
observations	6,478	2.306	13.795

Note: Data from children 10 - 16 years old in the wave 1 that were included in the estimation.

15 See http://www.ifpri.org/data/dataset.htm

Construction of Main Variables

The following is a rough indication on how the variables used in the analysis were constructed from the survey questions in wave 1. Expressions such as $p08$ refer to the number in the questionnaire and the variable name in the original data set.[16]

- S_i is the occupational choice variable. It takes value 0 if the answer to $p21$ is "No", it takes value 1 if the answer to $p21$ is "Yes" and any of the answers to $p22, p23, p301$ or $p302$ indicate that i works for income. It takes value 2 if the answer to $p21$ is "Yes" and the other aforementioned questions do not indicate labour income. If someone does not report income but reports to be working observations are dropped.
- Y_i is the weekly monetary income of i's household, i.e. $\sum_{i=1}^{N_i} y_i$
- y_i denotes the income of individual i. Using all sources of monetary income (questions $p22, p23, p301$ and $p302$), the payment received ($p291m, p31a2, p31b2$), the period of payment ($p291p, p31a1, p31b1$) and the amount of hours worked ($p2612$) a weekly income variable for each individual i is constructed.
- Y_{-i} is the household's income without i's contribution, i.e. $Y_i - y_i$.
- age_i refers to question $p08$.
- $female_i$ refers to question $p11$.
- edu_i refers to question $p20$. Years of schooling completed are calculated from *primaria* onwards, where each of the 6 *niveles* counts as one year. Hence after completion of the *secundaria* (3 *niveles*), one has 9 years of schooling.
- $rank_i$ is computed as the number of household members that are older than i but less than 19 years old plus one.
- N_i is the total number of persons living in the household.
- $child_i^{<6}$ is the total number of children below the age of 6 in the household of i.
- $poor_i$ indicates if i belongs to an eligible household, using *pobreden*.
- $treat_i$ indicates if i belongs to a treatment village, using *contbas2*.
- $state\_w_i$ the median earnings of i's gender/age group in his or her respective state. There are 7 states in the sample.

For the ex-post analysis, in addition to the aforementioned variables from wave 1, the analogous variables from waves 2 and 5 for age, schooling, gender etc. were used.

16 STATA codes are available upon request.

B.3 Simulation of Choice Consistent Residuals

For notational convenience, drop index i and call $x\beta$ the probability index that captures all variables. Then the underlying random utility model takes the form

$$U_j = x\beta_j + u_j.$$

Assuming that u_j derives from a double exponential distribution with independent draws, the choice probabilities p_j can be written as:[17]

$$p_j = \text{Prob}(x\beta_j + u_j > x\beta_k + u_k \text{ for all } k \neq j)$$

$$p_j = \frac{\exp^{x\beta_j}}{\sum_{k=0}^{2} \exp^{x\beta_k}}$$

The distribution of u_j conditional on the a choice $k \neq j$ has the following c.d.f.:

$$F(u_j \mid S_j = k) = \frac{\exp^{-\exp^{-u_j}}}{\exp^{-\exp^{x\beta_j - x\beta_k - u_k}}}.$$

The inverse of this function is used to draw residuals in the following sequence:

$$u_k = -\log(-p_k \cdot \log(rnd()) \qquad \text{if } j = k$$

$$u_j = -\log(\exp^{-u_k} \cdot (p_j/p_k)) - \log(rnd()) \qquad \text{if } j \neq k$$

where $rnd()$ is a random number between [0.1].

B.4 Bootstrap Mechanism Used

The algorithm consists of 4 steps.

Step 1 Compute the original sample estimates: estimation of equations (4.3) and (4.2) and simulation of the programme.

Step 2 Random draw with replacement of a new sample containing as many observations as the originial sample and compute all estimates described in step 1.

Step 3 Repeat step 2 for 1000 times collecting the predicted values of the age/gender specific impact. Use only the inner 90 percent of the values.

Step 4 Use the distribution of values obtained in step 3 to obtain confidence intervals around the estimates obtained in step 1.

17 For details, see Bourguignon et al. (2001a). I am grateful to Phillippe Leite for helpful comments on this point and sharing the STATA code.

B.5 Additional Tables

Table B.4: Pre-programme differences, boys

age	10	11	12	13	14	15	16
				full sample			
progresa	0.972	0.956	0.892	0.892	0.659	0.496	0.344
control	0.977	0.952	0.883	0.753	0.647	0.467	0.322
difference	-0.005	0.004	0.009	0.048	0.011	0.028	0.022
t-value	0.65	0.34	0.60	**2.33**	0.46	1.13	0.89
				not eligible			
progresa	0.975	0.985	0.906	0.864	0.772	0.622	0.298
control	0.975	0.961	0.924	0.773	0.712	0.54	0.253
difference	0.000	0.024	-0.018	0.091	0.061	0.083	0.045
t-value	0.01	1.10	0.56	**2.08**	1.14	1.47	0.99
				eligible			
progresa	0.971	0.952	0.888	0.788	0.635	0.47	0.361
control	0.977	0.951	0.874	0.749	0.634	0.447	0.35
difference	-0.006	0.000	0.014	0.040	0.001	0.023	0.011
t-value	0.70	0.04	0.85	1.72	0.05	0.81	0.38

Note: Table reports for each cell school enrolment ratios, the difference and the t-values. First wave only.

Table B.5: Pre-programme differences, girls

age	10	11	12	13	14	15	16
				full sample			
progresa	0.970	0.955	0.808	0.703	0.557	0.409	0.286
control	0.967	0.955	0.850	0.677	0.544	0.374	0.236
difference	0.003	0.000	-0.042	0.026	0.013	0.035	0.049
t-value	0.38	0.02	**2.20**	1.13	0.51	1.37	**2.01**
				not eligible			
progresa	0.972	0.974	0.867	0.720	0.669	0.523	0.298
control	0.990	0.956	0.894	0.788	0.589	0.377	0.222
difference	-0.018	0.018	-0.026	-0.068	0.080	0.146	0.076
t-value	1.00	0.78	0.67	1.30	1.41	**2.55**	1.57
				eligible			
progresa	0.970	0.952	0.795	0.700	0.531	0.378	0.282
control	0.963	0.955	0.841	0.653	0.533	0.373	0.242
difference	0.007	-0.003	-0.045	0.047	-0.002	0.005	0.040
t-value	0.75	0.25	**2.09**	1.80	0.07	0.19	1.41

Note: Table reports for each cell school enrolment ratios, the difference and the t-values. First wave only.

C Appendix to Chapter 5

C.1 Asymptotic Normality of LLC with Common Shocks

Absence of Shocks

Consider a model without constant and no additional lagged differences, along the lines of Levin et al. (2002). Following the notation from Section 5.2.1, in this case, for large N and T, no adjustments are necessary and $t_\delta^* = t_\delta$. The least squares estimator of δ proposed by LLC under the null hypothesis is:[18]

$$\hat{\delta} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}x_{it-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it-1}^2}$$

Define

$$\xi_{1iT} = \frac{1}{\sigma^2 T}\sum_{t=1}^{T}\epsilon_{it}x_{it-1} \text{ and } \xi_{2iT} = \frac{1}{\sigma^2 T^2}\sum_{t=1}^{T}x_{it-1}^2$$

and, using an estimator for the standard deviation[19] σ, the corresponding t−value is:

$$t_\delta = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{1iT}}{\left(\frac{\hat{\sigma}}{\sigma}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\xi_{2iT}\right]^{1/2}}.$$

Sectional correlation is a violation concerning the N. Taking the easiest form of multi index asymptotics, namely sequential limits (Phillips and Moon, 1999) first the limiting distributions when T goes to infinity is, for N fixed:[20]

$$\begin{aligned} \lim_{T\to\infty}\xi_{1iT} &= \xi_{1i} \text{ with } E[\xi_{1i}] = 0 \text{ and } Var[\xi_{1i}] = \frac{1}{2} \\ \lim_{T\to\infty}\xi_{2iT} &= \xi_{2i} \text{ with } E[\xi_{2i}] = \frac{1}{2} \text{ and } Var[\xi_{2i}] = \frac{1}{3} \end{aligned} \tag{C.1}$$

one obtains

$$t_\delta = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{1i}}{\left(\frac{\hat{\sigma}}{\sigma}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\xi_{2i}\right]^{1/2}}. \tag{C.2}$$

18 See Section 5.2.1.

19 For example:

$$\hat{\sigma} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}x_{it-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it-1}^2}.$$

20 These results are due to Phillips and Durlauf (1986), cf. Levin and Lin (1992, p. 14).

If the errors were uncorrelated and $\hat{\sigma}$ a consistent estimator of σ, averaging over the sections of the panel would give the known result that $t_{\delta} \Rightarrow N(0,1)$. The convergence in probability of the denominator of equation (C.2) is established by the following application of a law of large numbers (Billingsley, 1986, p. 290):

Theorem C.1 Suppose that for each time-series dimension T, the variables Z_{iT} are independent and identically distributed across individuals i, with mean μ_T and variance $0 < \sigma_T^2 < \infty$, and that $\mu = \lim_{T\to\infty} \mu_T$. If $\lim_{T\to\infty} \frac{\sigma_T^2}{N_T} = 0$. Then $\frac{1}{N_T} \sum_{i=1}^{N_T} Z_{iT} \xrightarrow{p} \mu$.

The inner expression of the denominator has expectation 1/2 and for all i the expectations of the variance are finite. Hence, the denominator converges to $\sqrt{0.5}$. The convergence in distribution of the numerator is established by applying the following central limit theorem (Billingsley, 1986, p. 368):

Theorem C.2 Suppose that for each time-series dimension T, the variables Z_{iT} are independent and identically distributed across individuals i, with mean μ_T and variance $0 < \sigma_T^2 < \infty$, and that $\mu = \lim_{T\to\infty} \mu_T$, and $\sigma^2 = \lim_{T\to\infty} \sigma_T^2$. Then $\frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} (Z_{iT} - \mu_T) \Rightarrow N(0, \sigma^2)$.

For each i, the numerator has expectation 0 and finite variance 1/2. Hence, it converges in distribution to $N(0, 0.5)$. Using the results obtained in (C.1) and applying both theorems, (C.2) converges to $N(0,1)$.[21]

Common Shocks

In the case of sectional correlation, however, the crucial assumption used in both theorems about the independence of the random variables is violated and their application fails. The numerator of (C.2) no longer converges to $N(0, 0.5)$. To be more precise, assume the easiest case in which the correlation among sections takes the following form:[22]

$$E[\xi_{1i}\xi_{1j}] \neq 0 \text{ and } Cov[\xi_{1i}\xi_{1j}] = \bar{\omega}_1 \text{ for } i \neq j$$

21 For the variance, notice that $Var(\frac{N(0,0.5)}{\sqrt{0.5}}) = \frac{0.5}{0.5} = 1$.

22 Note that if $Var(\epsilon_{it}) = 1$ the covariances equal the correlation coefficients.

To see which central limit theorem can be applied, it is necessary to check the properties of $\xi_1 \equiv \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{1i}$:

$$E[\xi_1] = E[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{1i}] = 0 \text{ and}$$

$$\begin{aligned} Var[\xi_1] &= Var[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{1i}] = \frac{1}{N}\Big(\sum_{i=1}^{N} Var(\xi_{1i}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(\xi_{1i}, \xi_{1j})\Big) \\ &= \frac{1}{N}\Big(\frac{N}{2} + N(N-1)\tilde{\omega}_1\Big) = \frac{1}{2} + \tilde{\omega}_1(N-1) \end{aligned}$$

The variance of the numerator increases with N. Central limit theorems for dependent random variables require a finite variance to establish convergence (see, *e.g.* Billingsley (1986, p. 376) and White (2001, p. 122)). Hence, no convergence result can be stated for this general form of dependence.

However, the analysis of the elimination of common time effects above (see page 107) has shown that the *effective* disturbance to the correlation matrix *after* removing common time effects is itself a function of N. More specifically, using the result from equation (5.6) that $\tilde{\omega}_1 = (1-\tilde{\omega}_1)(\frac{N-1}{N}\frac{-1}{N-1}) = \frac{\tilde{\omega}_1 - 1}{N}$ one can rewrite the above after removing common time effects as:

$$Var[\xi_1] = \frac{1}{2} + \frac{\tilde{\omega}_1 - 1}{N}.$$

As N goes to infinity the variance converges to the same value as in the case without sectional correlation. Using a central limit theorem which does not require independence (White, 2001, p. 125):

Theorem C.3 Suppose that for each time series dimension T, Z_{iT} is a stationary process with mean μ_T and variance $0 < \sigma_T^2 < \infty$, and that $Var(\frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} Z_{iT}) \xrightarrow{p} \sigma_N^2$ where $0 < \sigma_N^2 < \infty$. Then $\frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} (Z_{iT}) \Rightarrow N(0, \sigma_N^2)$.

Since each section has a finite variance and the variance of the average over all sections converges to $1/2$, the numerator converges to $N(0, 0.5)$. For the denominator write:

$$E[\xi_{2i}\xi_{2j}] \neq 0 \text{ and } Cov[\xi_{2i}\xi_{2j}] = \tilde{\omega}_2 \text{ for } i \neq j.$$

Again, checking the properties of $\xi_2 \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_{2i}$ it is easily seen that:

$$E[\xi_2] = \frac{1}{2} \text{ as before and}$$

$$\begin{aligned} Var[\xi_2] &= Var[\frac{1}{N}\sum_{i=1}^{N}\xi_{2i}] = \frac{1}{N^2}\Big(\sum_{i=1}^{N} Var(\xi_{2i}) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(\xi_{1i}, \xi_{1j})\Big) \\ &= \frac{1}{N^2}\Big(\frac{N}{3} + N(N-1)\tilde{\omega}_2\Big) = \frac{1}{3N} + \tilde{\omega}_2\frac{N-1}{N}. \end{aligned}$$

The variance of the denominator decreases with N. Following White (2001, p. 44), the following law of large numbers is applicable to weakly dependent data:

Theorem C.4 Suppose Z_i is a stationary ergodic scalar sequence with $E[Z_i] = \mu < \infty$. Then $Z_i \overset{a.s.}{\to} \mu$.

Almost sure convergence (a.s.) implies convergence in probability (Davidson (1994)). Hence, the denominator converges in probability to $\sqrt{(0.5)}$ regardless to the dependence in the data. Summarizing, with common sectional correlation and after removing common time effects, (C.2) will converge in distribution to $N(0, 1)$ as it the case without sectional correlation. This is in line with the simulation results obtained earlier, that the problem of oversizing diminishes with N and T.

Other Shocks

In the case where the covariance matrix takes the form of a band matrix (see page 109) if common time effects were not eliminated, the variance of the numerator, in terms of the expressions above, would again not converge for large N as it becomes $Var(\xi_1) = 1/2 + \sum_{a=1}^{N-1} a\omega^{N-a}$. If common time effects are eliminated, the structure of the correlation matrix becomes even less homogeneous and no convergence is achieved. The resulting matrix is a straightforward but rather unpleasant combination of N's and ω's. Here is a numerical example:

$$\text{if } \mathbf{\Sigma} = \begin{pmatrix} 1 & & & \\ .7 & 1 & & \\ .49 & .7 & 1 & \\ .34 & .49 & .7 & 1 \end{pmatrix} \text{ then } \mathbf{Q\Sigma Q'} = \begin{pmatrix} 1 & & & \\ .46 & 1 & & \\ -.81 & -.42 & 1 & \\ -.79 & -.81 & .46 & 1 \end{pmatrix}.$$

Hence, the elimination of common time effects in this case does not provide any remedy for the test. The test statistic will not converge to a N(0,1).

C.2 Estimation of and Inference on (Co-)Variances

Much attention has been devoted to so called heteroscedasticity and autocorrelation consistent (HAC) or robust estimators of covariances from to stationary series. Under nonstationarity, covariances are not constant over time and methods designed for stationary series can no longer be used. In this case one can either use the differences of the series to compute the covariances or the residuals from regression on the lagged variable. Schwert (1989) finds that, in the univariate case, the difference based approach has a smaller bias in finite samples. Therefore, and in order to proceed consistently under the null of non-stationarity, the following lines apply to the first differences of an I(1) process without drift.

Analogous to the univariate problem of variance estimation, the aim is to get a consistent estimate of the covariance matrix at zero frequency.[23] To estimate the spectrum of an unknown DGP correctly, all T autocovariances have to be estimated, which is not feasible with T observations. The class of parametric estimators imposes a certain structure on the data and constructs estimators that would be implied by the model, while nonparametric procedures use a weighted average of autocovariances.

Parametric Estimators

The parametric estimator VARHAC (vector autoregressive heteroscedasticity and autocorrelation consistent) was developed by Den Haan and Levin (1996) and fits a vector autoregressive (VAR) model to the series under consideration using an information criterion to determine the optimal lag length. To the residuals of that VAR a standard covariance estimator is applied.

More specifically, for each section i of the N-dimensional stationary of the process $\mathbf{x}_t$, an autoregressive process is fitted using a lag order suggested by either the Akaike (AIC) or the Schwarz' Bayesian (BIC) information criterion, and given a maximum lag order. The optimal lag order may differ across sections. The coefficients are collected in a matrix $\dot{\mathbf{A}}_{k(N\times N)}$ for each lag k, taking zero values for section i if k exceeds the maximum lag order of that section. For the highest lag length $\bar{K}$ chosen, a VAR is fitted and the residuals $\dot{\mathbf{e}}_t$ are used to compute the innovation matrix:

$$\dot{\Sigma}_T^{VARHAC} = \frac{1}{T}\sum_{t=\bar{K}}^{T} \dot{\mathbf{e}}_t\dot{\mathbf{e}}_t'.$$

23 For the following, see Den Haan and Levin (1997).

The the spectral density estimator is then given by:

$$\hat{\mathbf{S}}_T^{VARHAC} = [I - \sum_{k=1}^{\bar{K}} \hat{\mathbf{A}}_k]^{-1} \hat{\Sigma}^{VARHAC} [I - \sum_{k=1}^{\bar{K}} \hat{\mathbf{A}}_k]^{-1}. \tag{C.3}$$

Den Haan and Levin (1996) analyze the performance of this estimator compared to some nonparametric alternatives and find better finite sample properties. According to their results, the individual choice of lag lengths for each section makes this procedure superior to nonparametric estimates, in which one weighting function is applied to all sections.

Nonparametric Estimators

In the nonparametric case two concepts are introduced to handle the problem of estimating the covariances: windowing and weighting. The most frequently used kernels in the time series literature are the Bartlett kernel and the Parzen kernel. For the Bartlett kernel, the weights assigned to the autocovariances decline from 1 (the sample variance) to 0 (when the truncation is reached). This kernel ensures a positive estimation of the long-run variance – or, in the multivariate case – a positive definite estimate of the covariance matrix (Newey and West, 1987). Since the theoretical guidance in the choice of the truncation is quite unsatisfactory, it might be useful in empirical applications to conduct robustness checks in terms of varying kernels and truncation parameters. Starting point for the estimation of the covariance matrix in the presence of serial correlation is:

$$\mathbf{S}(m) = \hat{\mathbf{\Gamma}}_0 + \sum_{\tau=1}^{K} w_{K\tau} (\hat{\mathbf{\Gamma}}_\tau + \hat{\mathbf{\Gamma}}'_\tau), \tag{C.4}$$

where $w_{K\tau}$ is a kernel, K a truncation parameter, and

$$\hat{\mathbf{\Gamma}}_\tau = \frac{1}{T} \sum_{t=\tau+1}^{T} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t-\tau} - \bar{\mathbf{x}})'.$$

Refinements to this estimator are possible. Kernel based estimations of the long-run variance matrix in the presence of serial correlation were found to give quite poor results. The major source of bias is that kernels, which ensure a positive definite spectral density matrix place weights less then unity on autocovariances other then at lags zero. Andrews and Monahan (1992) therefore suggest a kernel based prewhitening of the series and observe an improvement using this technique. In an expression similar to equation (C.3) the covariance matrix is placed between the inverse of the prewhitening coefficients. Newey and West (1994) propose an automated bandwidth selection procedure for the estimator in equation (C.4).

C.3 Bootstrap Methods for Dependent Data

Inference on covariance matrix estimators is rarely done. But the estimation results itself are meaningless if they remain unrelated to some standard errors. In both the parametric and nonparametric case, bootstrap methods may be used to make inference on the estimates.

There is little known about the properties of bootstrap algorithms when the underlying process contains a unit root. But even if the root of the process comes close to unity, Bose (1988) shows that bootstrap approximations deteriorate. However, bootstrapping results will remain valid if the bootstraps are applied to the differenced data. Hence, the following applies to the first differences of a non-stationary process.

For time dependent data, however, the algorithms have to be extended because random resampling would not account for the time dependency of the data, which, as it is the case for the covariance matrix, is a crucial part of the estimator.[24] For time dependent processes, resampling in the frequency domain is suggested *e.g.* by Franke and Haerdle (1992) for the univariate case. This method is designed for making inference about the entire spectrum. In this context the only estimate of interest is the variation of the covariance matrix at zero frequency, and therefore these methods do not seem appropriate. In the time series domain the following methods are suggested. The so called model-based resampling requires reasonable good knowledge of the true model. In short, the assumed DGP is applied to the series, innovations are computed and then used to resample a series again assuming the same DGP. Among the methods that do not require knowledge of the DGP is the so called block resampling. The basic idea here is to divide the data into b blocks of equal length l. The new resampled series is a randomly order of blocks. Typically, those estimates will be biased, because the resampled series are more independent than the original one, since whenever a block changes, artificial independence is introduced. Furthermore, this break causes the artificial series to exhibit nonstationarity properties, because distribution parameters become time dependent.[25]

The Stationary Bootstrap

The stationary bootstrap suggested by Politis and Romano (1994) is a sophistication of the aforementioned methods. Moreover, this bootstrap is unbiased and does not produce nonstationarity in the above sense. Another advantage of this method is that its validity for the covariance estimation of a multi-

24 See *e.g.* Davison and Hinkley (1997).

25 For methods on how to overcome this and other problems, see Hall et al. (1995).

variate process was shown, which is precisely what is needed in this context (Politis and Romano, 1994, Theorem 4). The algorithm is as follows:

- Let $\mathbf{x}_t$ be a N-dimensional vector of time series from $t = 1, ..., T$.
- Define $b_{t,l} = \{\mathbf{x}_t, \mathbf{x}_{t+1}, ...\mathbf{x}_{t+l-1}\}$ as a block of l subsequent observations in the sample, starting at some t. If the end of the sample is reached before the end of the block (*i.e.* $t + l > T$), the block is filled up with observations of the beginning of the sample ($\mathbf{x}_N = \mathbf{x}_0$, $\mathbf{x}_{N+1} = \mathbf{x}_1$ ...).
- The length l of the blocks is determined randomly, where the lengths follow a geometrical distribution with some fixed parameter $p \in [0, 1]$. The probability of a block length m is $\Pr\{l = m\} = (1-p)^{(m-1)}p$ for $m = 1, 2, ...$. Denote those random numbers by L_i.
- Once the lag length is determined, the beginning of the block is determined by a random variable I_i which is discretely uniformly distributed on $[1, T]$.
- The pseudo time series $\mathbf{x}_t^* = \{\mathbf{x}_1^*...\mathbf{x}_T^*\}$ is generated by the random sequence of blocks $B_{I_1,L_1}, B_{I_2,L_2}...$, where the end is trimmed at T. The resampling is done B times.
- Let the true vector of parameters of interest be $\boldsymbol{\theta}$. In the same way as the distribution of $\mathbf{x}$ can be approximated by the large number of pseudo series $\mathbf{x}^*$, the distribution of $\dot{\boldsymbol{\theta}}$ conditional on $\mathbf{x}$ can be approximated by the distribution of $\dot{\boldsymbol{\theta}}^*(\mathbf{x}^*)$.
- Applying this procedure to the inference on a covariance matrix, $\boldsymbol{\theta}$ is a vector containing the correlations between the N units of the the vector $\mathbf{x}_t$. If one restricts attention to the triangle below the diagonal, this amounts to $(\frac{1}{2}(N-1)N) = d$ elements. Denote by $\dot{\boldsymbol{\theta}}(\mathbf{x})$ a consistent (parametric or nonparametric) estimator of the covariance matrix. After computing the covariances for each resampled $\mathbf{x}^*$, one can estimate the asymptotic variance of the estimator by:[26]

$$\dot{\mathbf{V}}[\dot{\boldsymbol{\theta}}] = \frac{1}{B}\sum_{b=1}^{B}[\dot{\boldsymbol{\theta}}^*(b) - \dot{\boldsymbol{\theta}}][\dot{\boldsymbol{\theta}}^*(b) - \dot{\boldsymbol{\theta}}]'. \tag{C.5}$$

The diagonal elements of $\dot{\mathbf{V}}$ contain the variances of each element of the estimator, hence the root of the diagonal contains the standard error to be placed around the point estimates $\dot{\boldsymbol{\theta}}$.

26 See Greene (2000, p. 174).

- Assuming that the bootstrapped values follow a normal distribution, then a simple test for $\dot{\boldsymbol{\theta}} = \boldsymbol{\theta}_1$ is:

$$Q = (\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}_1)'\hat{\mathbf{V}}^{-1}(\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}_1) \tag{C.6}$$

where $\hat{\mathbf{V}}$ is an estimate of the covariance matrix of the form:

$$\hat{\mathbf{V}}[\dot{\boldsymbol{\theta}}] = \frac{1}{B}\sum_{b=1}^{B}[\dot{\boldsymbol{\theta}}^*(b)][\dot{\boldsymbol{\theta}}^*(b)]'$$

Then Q will be approximately χ^2_d distributed, where d is the dimension of $\boldsymbol{\theta}$.[27]

C.4 Additional Tables and Figures

Table C.1: Size properties when all assumptions are fulfilled

$\omega = .0$	LLC			IPS		
	nominal size 10%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.136	.120	.118	.120	.114	.109
$T = 50$	.122	.101	.101	.114	.100	.104
$T = 100$	.118	.100	.100	.101	.101	.098
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.073	.058	.058	.069	.059	.059
$T = 50$	.062	.053	.047	.064	.052	.054
$T = 100$	.058	.056	.049	.058	.057	.048
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.015	.011	.010	.018	.012	.012
$T = 50$	.013	.011	.009	.015	.013	.009
$T = 100$	.013	.010	.009	.013	.012	.009

Note: Based on 4,000 replications.

27 See also Den Haan and Levin (1997, p. 299).

Table C.2: Size properties with cointegration

$CIV = N/2$	LLC			IPS		
	nominal size 10%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.155	.146	.147	.150	.142	.142
$T = 50$	.169	.151	.145	.179	.185	.194
$T = 100$	.178	.160	.149	.267	.291	.301
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.081	.080	.077	.081	.083	.080
$T = 50$	.086	.077	.072	.099	.103	.116
$T = 100$	.010	.088	.078	.167	.188	.202
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.017	.014	.016	.020	.017	.022
$T = 50$	.020	.014	.013	.031	.027	.024
$T = 100$	.026	.022	.014	.052	.064	.072

Note: Based on 4,000 replications.

Table C.3: Size properties with cointegration, $b = N/4$

$b = N/4$	LLC			IPS		
	nominal size 10%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.144	.137	.135	.126	.132	.139
$T = 50$	.147	.132	.130	.152	.143	.142
$T = 100$	.164	.146	.142	.184	.184	.186
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.074	.072	.073	.069	.077	.074
$T = 50$	.074	.070	.072	.085	.078	.077
$T = 100$	.070	.082	.074	.108	.109	.112
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.016	.014	.014	.019	.021	.017
$T = 50$	.014	.016	.017	.021	.016	.019
$T = 100$	.021	.017	.019	.029	.026	.024

Note: Based on 4,000 replications.

Table C.4: Size properties with cointegration and correlation for $b = N - 1$

$b = N-1$	LLC			IPS		
	nominal size 10%					
$\omega = 0.7$	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.092	.056	.023	.079	.045	.022
$T = 50$	.084	.045	.020	.086	.032	.013
$T = 100$	.052	.038	.013	.122	.062	.019
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.045	.027	.017	.042	.021	.011
$T = 50$	.042	.019	.007	.043	.013	.005
$T = 100$	.025	.018	.007	.064	.031	.008
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.010	.007	.003	.011	.006	.002
$T = 50$	.011	.004	.001	.010	.003	.001
$T = 100$	.008	.005	.001	.018	.007	.001

Note: Based on 4,000 replications.

Table C.5: Size properties with cointegration and correlation for $b = N/4$

$b = N/4$	LLC			IPS		
	nominal size 10%					
$\omega = 0.7$	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.080	.054	.026	.059	.033	.013
$T = 50$	.091	.068	.029	.046	.021	.005
$T = 100$	.118	.068	.039	.048	.016	.005
	nominal size 5%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.043	.029	.012	.034	.017	.005
$T = 50$	.054	.041	.014	.023	.011	.003
$T = 100$	.081	.042	.025	.028	.010	.001
	nominal size 1%					
	$N = 5$	$N = 10$	$N = 20$	$N = 5$	$N = 10$	$N = 20$
$T = 25$	.013	.007	.002	.008	.003	.001
$T = 50$	.017	.012	.004	.005	.002	.001
$T = 100$	.030	.018	.009	.008	.003	.000

Note: Based on 4,000 replications.

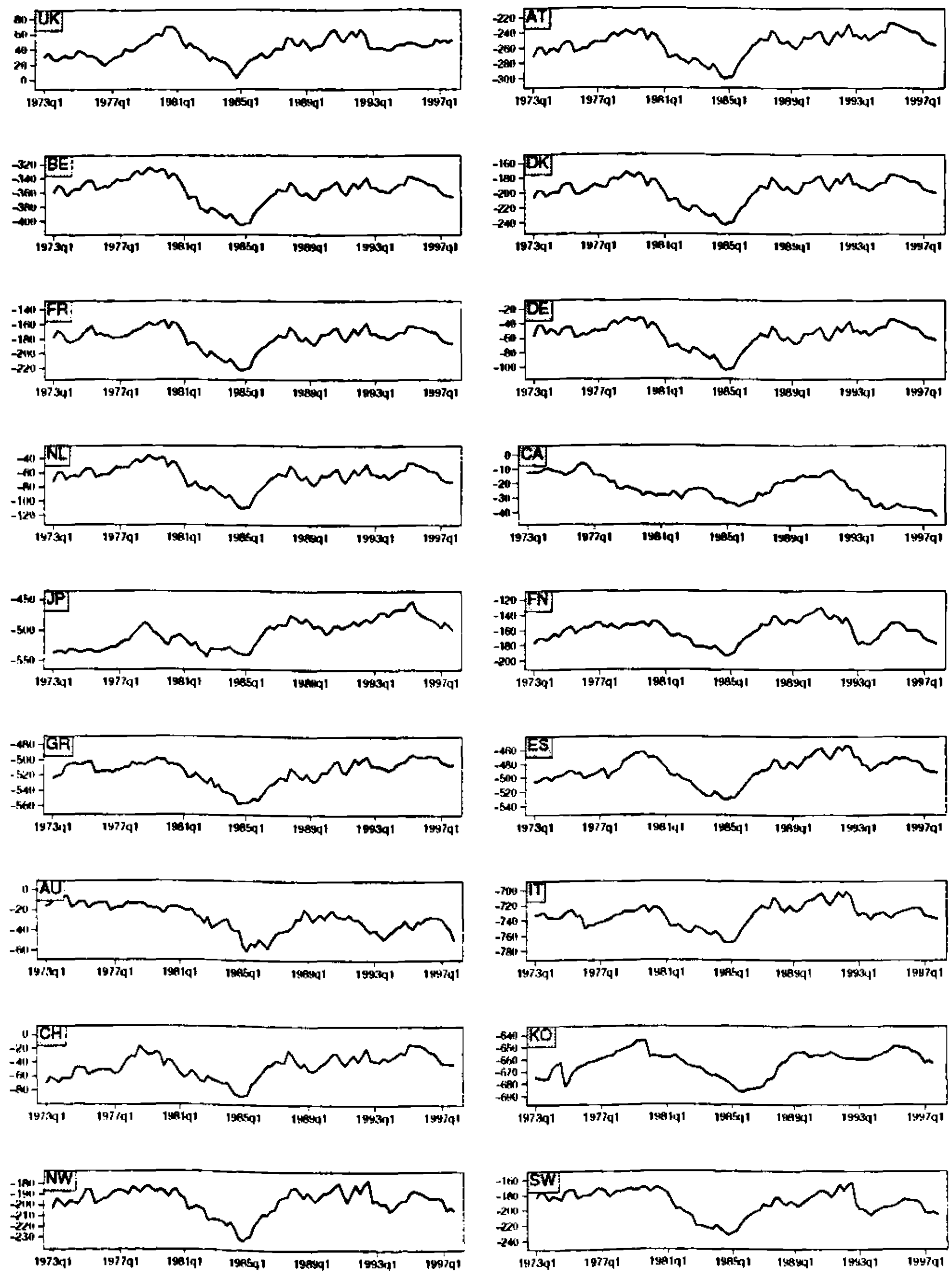


Figure C.1: Exchange rates with US$ as base currency

Table C.6: Estimated sectional correlation in the differenced exchange rates

	UK	AT	BE	DK	FR	DE	NL	CA	JP	FN	GR	ES	AU	IT	CH	KO	NW	SW
UK	1.00																	
AT	0.60 0.10	1.00																
BE	0.64 0.10	0.97 0.02	1.00															
DK	0.62 0.10	0.98 0.01	0.98 0.01	1.00														
FR	0.72 0.09	0.92 0.04	0.91 0.03	0.93 0.04	1.00													
DE	0.60 0.10	0.99 0.01	0.95 0.03	0.96 0.02	0.92 0.04	1.00												
NL	0.61 0.10	0.99 0.01	0.97 0.02	0.97 0.01	0.92 0.04	0.99 0.01	1.00											
CA	−0.01 0.18	0.03 0.15	0.00 0.15	0.07 0.15	−0.06 0.15	0.02 0.16	0.01 0.16	1.00										
JP	0.32 0.17	0.61 0.13	0.61 0.13	0.59 0.13	0.53 0.15	0.60 0.13	0.61 0.13	−0.05 0.15	1.00									
FN	0.65 0.10	0.75 0.10	0.75 0.10	0.77 0.09	0.73 0.10	0.71 0.11	0.75 0.10	0.25 0.19	0.36 0.18	1.00								
GR	0.64 0.11	0.83 0.09	0.80 0.10	0.83 0.09	0.76 0.13	0.82 0.10	0.83 0.10	0.14 0.16	0.41 0.13	0.70 0.09	1.00							
ES	0.68 0.10	0.81 0.08	0.82 0.08	0.84 0.08	0.83 0.08	0.79 0.09	0.81 0.08	0.09 0.19	0.35 0.18	0.78 0.08	0.79 0.10	1.00						
AU	0.30 0.14	0.26 0.13	0.21 0.13	0.26 0.13	0.25 0.13	0.25 0.14	0.25 0.14	0.41 0.17	0.25 0.14	0.44 0.14	0.39 0.12	0.33 0.12	1.00					
IT	0.73 0.11	0.79 0.09	0.79 0.09	0.80 0.09	0.84 0.07	0.78 0.10	0.78 0.09	0.09 0.19	0.37 0.20	0.70 0.13	0.73 0.11	0.81 0.08	0.32 0.13	1.00				
CH	0.60 0.12	0.91 0.04	0.90 0.05	0.91 0.04	0.88 0.04	0.90 0.04	0.91 0.04	0.02 0.16	0.65 0.12	0.69 0.11	0.77 0.09	0.73 0.11	0.24 0.14	0.71 0.10	1.00			
KO	0.02 0.13	0.18 0.19	0.17 0.17	0.17 0.19	0.13 0.18	0.17 0.22	0.16 0.20	0.30 0.15	0.24 0.12	0.24 0.15	0.22 0.18	0.23 0.14	0.53 0.20	0.09 0.15	0.15 0.18	1.00		
NW	0.71 0.09	0.89 0.03	0.88 0.04	0.90 0.04	0.86 0.05	0.87 0.04	0.89 0.04	0.17 0.19	0.51 0.13	0.86 0.06	0.80 0.09	0.81 0.09	0.41 0.14	0.77 0.10	0.81 0.06	0.19 0.19	1.00	
SW	0.71 0.11	0.76 0.08	0.80 0.07	0.80 0.07	0.80 0.08	0.73 0.08	0.76 0.07	0.15 0.19	0.36 0.18	0.83 0.07	0.70 0.12	0.86 0.07	0.39 0.14	0.81 0.10	0.67 0.09	0.15 0.16	0.86 0.06	1.00
	UK	AT	BE	DK	FR	DE	NL	CA	JP	FN	GR	ES	AU	IT	CH	KO	NW	SW

Note: The values reported are correlation coefficients obtained after the correlation transformation of the estimator in equation (C.4) using $K = 3$. The standard errors reported are based on 500 replications of the bootstrap procedure described in Section C.3 and computed as in equation (C.5).

List of Tables

List of Figures

Bibliography

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Anker, R. (2000). *Conceptual and Research Frameworks for the Economics of Child Labour and its Elimination.* IPEC Working Paper. Geneva: International Labour Organization.

Attanasio, O., C. Meghir, and A. Santiago (2001). *Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA.* Mimeo. University College London.

Banerjee, A., M. Marcellino, and C. Osbat (2000). *Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data.* EUI Working Papers (2000/20). Florence: European University Intitute.

Banerjee, A., M. Marcellino, and C. Osbat (2001). *Testing for PPP: Should We Use Panel Methods.* Mimeo. European University Institute.

Behrman, J., P. Sengupta, and P. Todd (2001). *Progressing Through PROGRESA: An Impact Assessment of a School Subsidy Experiment.* Pier Working Paper No. 01-033. Philadelphia: Penn Institute for Economic Research.

Bentolila, S. and A. Ichino (2003). *Unemployment and Consumption: Why are Job Losses Less Painful Than Expected Near the Mediterranean?* mimeo: European University Institute.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity and social history. *Games and Economic Behaviour 10*, 122–142.

Billingsley, P. (1986). *Probability and Measure* (2nd ed.). Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Bornhorst, F. and C. Baum (2001a). ***ipshin.ado, ipshin.hlp**: STATA Module to Perform the Im et al. (1997) Panel Unit Root Test.* Available at http://ideas.repec.org/c/boc/bocode/s419704.html.

Bornhorst, F. and C. Baum (2001b). ***levinlin.ado, levinlin.hlp**: STATA Module to Perform the Levin et al. (2002) Panel Unit Root Test.* Available at http://ideas.repec.org/c/boc/bocode/s419702.html.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics 16*, 1709–1722.

Bourguignon, F., F. H. G. Ferreira, and P. Leite (2002). *Ex-Ante Evaluation of Conditional Cash Transfer Programs: The Case of Bolsa Escola.* World Bank Policy Research Working Paper Nr. 2916. Washington, DC: The World Bank.

Bourguignon, F., F. H. G. Ferreira, and P. G. Leite (2003). Conditional cash transfers, schooling, and child labor: Micro-simulating Brazil's Bolsa Escola program. *World Bank Economic Review 17*(2), 229–254.

Bourguignon, F., M. Fournier, and M. Gurgand (2001a). Fast development with a stable income distribution: Taiwan, 1979 - 1994. *Review of Income and Wealth 47*(2), 139–163.

Bourguignon, F., M. Fournier, and M. Gurgand (2001b). *Selection Bias Correction Based on the Multinomial Logit Model.* Mimeo. World Bank and DELTA.

Breitung, J. and W. Meyer (1991). *Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Constrained?* Institut fuer Quantitative Wirtschaftsforschung Working Paper. Humboldt Universitaet zu Berlin.

Buchan, N., R. T. Croson, and E. J. Johnson (2000). *Trust and Reciprocity: An International Experiment.* Mimeo. University of Wisconsin.

Buchan, N. R., S. Solnick, and R. T. Croson (2004). *Trust and Gender: An Examination of Behavior, Biases, and Beliefs in the Investment Game.* Mimeo. Pennsylvania: The Wharton School of the University of Pennsylvania.

Burks, S. V., J. P. Carpenter, and E. Verhoogen (2003). Playing both roles in the trust game. *Journal of Economic Behaviour and Organization 51*(2), 195–216.

Cheung, Y. W. and K. S. Lai (1993). Long run purchasing power parity during the recent float. *Journal of International Economics 34*, 181–192.

CIA (2003). *The World Fact Book*. available online at http://www.cia.gov/cia/publications/factbook/.

Cochard, F., N. V. Phu, and M. Willinger (2000). *Trust and Reciprocity in a Repeated Investment Game*. Mimeo. Laboratoire d'Economie Expermientale de Strasbourg.

Corbae, D. and Ouliaris (1991). A test of long run PPP allowing for structural breaks. *Economic Record 67*, 26–33.

Cox, J. C. (2002). *How to Identify Trust and Reciprocity*. Mimeo. University of Arizona.

Croson, R. T. and N. Buchan (1999). Gender and culture: International experimental evidence from a trust game. *American Economic Review: Papers and Proceedings 89*, 386–391.

Crowder, W. J. (1997). *Purchasing Power Parity During the Modern Float: A Look Behind The Panels*. Working Paper. University of Texas at Arlington.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford: Oxford University Press.

Davison, A. and D. Hinkley (1997). *Bootstrap Methods and their Applications*. Cambridge: Cambridge University Press.

Den Haan, W. J. and A. Levin (1996). *Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures*. NBER Technical Working Paper Series 195. Cambridge, MA: NBER.

Den Haan, W. J. and A. T. Levin (1997). A practitioner's guide to robust covariance matrix estimation. In G. Maddala and R. C.R. (Eds.), *Handbook of Statistics*, Volume 15, Chapter 12, pp. 299–342. Amsterdam: Elsevier.

Dickey, D. and W. Fuller (1981). Likelihood ratio tests for autoregressive time series with a unit root. *Econometrica 49*, 1057–1072.

Eckel, C. C. and R. K. Wilson (2003). *Conditional Trust: Sex, Race and Facial Expressions in a Trust Game.* Mimeo. Blacksburg, VA: Virginia Tech.

Economist (2001, April 21st). *Economics Focus: The Big Mac Index.* London.

Edison, H. J., J. Gagnon, and W. Melick (1997). Understanding the empirical literature on purchasing power parity: The post bretton woods era. *Journal of International Money and Finance 16*, 1–17.

El-Gamal, M. A. and D. M. Grether (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American Statistical Association 90*(432), 1137–1145.

Engel, C. (2000). Long-run PPP may not hold after all. *Journal of International Economics 57*, 243–273.

Erev, I. and A. Roth (1998). Predicting how poeple play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review 88*(4), 848–881.

Evans, P. and G. Karras (1996). Do economies converge? evidence from a panel of U.S. states. *Review of Economics and Statistics 78*(3), 384–388.

Falk, A. (2003). Homo Oeconomicus versus Homo Reciprocans: Ansaetze fuer ein neues Wirtschaftspolitisches Leitbild? *Perspektiven der Wirtschaftspolitik 4*(1), 141–172.

Falk, A. and U. Fischbacher (1999). *A Theory of Reciprocity.* Working Paper No. 6. University of Zurich: Institute for Empirical Research in Economics.

Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *Quaterly Journal of Economics 116*(1), 351–377.

Fischbacher, U. (1999). *Z-Tree: Zurich Toolbox for Readymade Economic Experiments.* Working Paper No. 21. University of Zurich: Institute for Empirical Research in Economics.

Franke, J. and W. Haerdle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics 20*(1), 121–145.

Froot, K. and K. Rogoff (1995). Perspectives of PPP and long-run real exchange rates. In G. Grossman and K. Rogoff (Eds.). *Handbook of International Economics*. Amsterdam: North-Holland.

Glaeser, E., D. L. Laibson, J. A. Scheinkman, and C. L. Soutter (2000). Measuring trust. *Quaterly Journal of Economics 115*(3), 811–846.

Greene, W. H. (2000). *Econometric Analysis* (4^{th} ed.). New Jersey: Prentice Hall.

Guiso, L., P. Sapienza, and L. Zingales (2003). *Cultural Biases in Economic Exchange*. Mimeo. Northwestern University.

Hall, P., J. L. Horowitz, and B. Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*(3), 561–574.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*, 153–161.

ILO (2004). *Investing in Every Child: An Economic Study of the Costs and Benefits of Eliminating Child Labour*. Geneva: International Labour Organization.

Im, K., M. Pesaran, and Y. Shin (1997). *Testing for Unit Roots in Heterogeneous Panels*. Discussion Paper 9526, Department of Applied Economics, University of Cambridge.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford and New York: Oxford University Press.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decisions under risk. *Econometrica 47*, 313–327.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *Quarterly Journal of Economics 52*, 1251–1287.

Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology 100*, 313–345.

Kouretas, G. P. (1997). The canadian dollar and purchasing power parity during the recent float. *Review of International Economics 5*(4), 467–477.

Kuo, B. S. and A. Mikkola (1999). *How Sure are We About PPP? Panel Evidence With the Null of Stationary Real Exchange Rates.* CEPR Discussion Paper: 2120. CEPR.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*(4), 604–620.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica 51*, 507–512.

Levin, A. and C.-F. Lin (1992). *Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Porpcrties.* University of California Discussion Paper (92-23). University of California, San Diego.

Levin, A. and C.-F. Lin (1993). *Unit Root Tests in Panel Data: New Results.* University of California Discussion Paper (93-56). University of California, San Diego.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics 108*(1), 1–24.

Lyhagen, J. (2000). *Why Not Use Standard Panel Unit Root Test for Testing PPP.* Volume 413 of *Stockholm School of Economics Working Paper Series in Economics and Finance.* Stockholm School of Economics.

Maddala, G. S. (1993). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Manski, C. F. (2002). Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review 46*, 880–891.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics.* New York: Academic Press.

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: Columbia University Press.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*, 385–397.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*, 631–653.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.

O'Connell, P. G. (1998). The overvaluation of purchasing power parity. *Journal of International Economics 44*, 1–19.

OECD (2003). *Combating Child Labour: A Review of Policies.* Paris: Organisation for Economic Cooperation and Development.

Papell, D. (1997). Searching for stationarity: Purchasing power parity under the current float. *Journal of International Economics 43*, 313–323.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics 61*, 653–670.

Pesaran, M. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Phillips, P. and S. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies 53*, 473–495.

Phillips, P. C. B. and H. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*, 1057–1111.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Quah, D. (1990). *International Patterns of Growth: I. Persistence in Cross-Country Disparities.* MIT Working Paper.

Schultz, P. T. (2001). *School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program.* Economic Growth Center Discussion Paper Nr. 834. New Haven: Yale University.

Schwert, G. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business and Economics Statistics 7*, 147–159.

Selten, R. and R. Stoecker (1986). End behaviour in sequences of finite prisoner's dilemma supergames - a learning theory approach. *Journal of Economic Behavior and Organization 7*(1), 47–70.

Strauss, J. (2000). Is there a permanent component in US real GDP? *Economics Letters 66*(2), 137–142.

Todd, P. and K. I. Wolpin (2003). *Using a Social Experiment to Validate a Dynamic Behavioural Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico.* Mimeo. University of Pennsylvania.

White, H. (2001). *Asymptotic Theory for Econometricians* (2nd ed.). London: Academic Press.

Worldbank (2000). *World Development Report: Attacking Poverty.* Washington DC: World Bank.

Wu, J. L. (2000). Mean reversion of the current account: Evidence from the panel data unit-root test. *Economics Letters 66*(2), 215–222.

L1
30
.1
B